Journal of
Educational Technology & Society

Published by International Forum of Educational Technology & Society
Hosted by National Yunlin University of Science and Technology, Taiwan

Volume 24 Issue 1
January 2021
ISSN: 1436-4522 (online)
ISSN: 1176-3647 (print)
http://www.j-ets.net/

Jan. 2021 Journal of Educational Technology & Society
vol.24
no.1

Educational Technology & Society has Impact Factor 2.086 and
5-Year impact factor of 2.720 according to Thomson Scientific 2019 Journal Citations Report.

http://www.j-ets.net/

Educational Technology & Society

An International Journal

Aims and Scope

The journal of *Educational Technology & Society* (ET&S) is an open-access academic journal published quarterly (January, April, July, and October) since October 1998. By 2018, ET&S has achieved its purposes at the first stage by providing an international forum for open access scientific dialogue for developers, educators and researchers to foster the development of research in educational technology. Thanks to all of the Authors, Reviewers and Readers, the journal has enjoyed tremendous success.

Starting from 2019, the ET&S journal has established a solid and stable editorial office with the support of National Yunlin University of Science and Technology. The new Editors-in-Chief have been appointed aiming to promote innovative educational technology research based on empirical inquires to echo the pedagogical essentials of learning in the real world—lifelong learning, competency-orientation, and multimodal literacy in the 21st century.

ET&S publishes the research that well bridges the pedagogy and practice in advanced technology for evidence-based and meaningfully educational application. The focus of ET&S is not only technology per se, but rather issues related to the process continuum of learning, teaching, and assessment and how they are affected or enhanced using technologies rooted in a long-period base. The empirical research about how technology can be used to overcome the existing problems in the frontline of local education with findings that can be applied to the global spectrum is also welcome. However, papers with only descriptions of the results obtained from one hit-and-run and short-term study or those with the results obtained from self-report surveys without systematic or empirical data or any analysis on learning outcomes or processes are not favorable to be included in ET&S.

Publisher

International Forum of Educational Technology & Society

Host

National Yunlin University of Science and Technology, Taiwan

Editorial Office

c/o Chair Professor Nian-Shing Chen, National Yunlin University of Science and Technology, No. 123, Section 3, Daxue Road, Douliu City, Yunlin County, 64002, Taiwan.

Supporting Organizations

University of North Texas, USA
University of Piraeus, Greece

Advertisements

Educational Technology & Society accepts advertisement of products and services of direct interest and usefulness to the readers of the journal, those involved in education and educational technology. Contact the editors at journal.ets@gmail.com

Abstracting and Indexing

Educational Technology & Society is abstracted/indexed in Social Science Citation Index, Current Contents/Social & Behavioral Sciences, ISI Alerting Services, Social Scisearch, ACM Guide to Computing Literature, Australian DEST Register of Refereed Journals, Computing Reviews, DBLP, Educational Administration Abstracts, Educational Research Abstracts, Educational Technology Abstracts, Elsevier Bibliographic Databases, ERIC, JSTOR, Inspec, Technical Education & Training Abstracts, and VOCED.

Guidelines for authors

Submissions are invited in the following categories:

- Peer reviewed publications: Full length articles (4,000 to 8,000 words)
- Special Issue publications

All peer review publications will be refereed in double-blind review process by at least two international reviewers with expertise in the relevant subject area.

For detailed information on how to format your submissions, please see:
https://www.j-ets.net/author_guide

For Special Issue Proposal submission, please see:
https://www.j-ets.net/journal_info/special-issue-proposals

Submission procedure

All submissions must be uploaded through our online management system (http://www.j-ets.net). Do note that all manuscripts must comply with requirements stated in the Authors Guidelines.

Authors, submitting articles for a particular special issue, should send their submissions directly to the appropriate Guest Editor. Guest Editors will advise the authors regarding submission procedure for the final version.

All submissions should be in electronic form. Authors will receive an email acknowledgement of their submission.

The preferred formats for submission are Word document, and not in any other word-processing or desktop-publishing formats. For figures, GIF and JPEG (JPG) are the preferred formats. **Authors must supply separate figures** in one of these formats besides embedding in text.

Please provide following details with each submission in a separate file (i.e., Title Page): ▪ Author(s) full name(s) including title(s), ▪ Name of corresponding author, ▪ Job title(s), ▪ Organisation(s), ▪ Full contact details of ALL authors including email address, postal address, telephone and fax numbers.

In case of difficulties, please contact **journal.ets@gmail.com** (Subject: Submission for Educational Technology & Society journal).

Educational Technology & Society
Volume 24 Number 1 2021

Table of contents

Full Length Articles

Learning Analytics for Collaborative Language Learning in Classrooms: From the Holistic Perspective of Learning Analytics, Learning Design and Teacher Inquiry — 1–15
Yun Wen and Yanjie Song

A Multiple Study Investigation of the Evaluation Framework for Learning Analytics: Instrument Validation and the Impact on Learner Performance — 16–28
Mehmet Kokoç and Mehmet Kara

Can Public Health Workforce Competency and Capacity be built through an Agent-based Online, Personalized Intelligent Tutoring System? — 29–43
Sarah D. Matthews and Michael D. Proctor

Using Gamification to Design Courses: Lessons Learned in a Three-year Design-based Study — 44–63
Huang Biyun and Hwe Khe Foon

Organizing and Hosting Virtual PPTELL 2020 During the COVID-19 Pandemic — 64–74
Yu-Ju Lan, Kao Chia-Ling Gupta, Tai-Yi Huang, Shobhana Chelliah and J. Michael Spector

Using Data Analytics to Investigate Attendees' Behaviors and Psychological States in a Virtual Academic Conference — 75–91
Jiun-Yu Wu, Chen-Hsuan Liao, Tzuying Cheng and Mei-Wen Nian

Challenges in Organizing Online Conferences: Lessons of the COVID-19 Era — 92–104
Margus Pedaste and Meriliis Kasemets

Editorial Note

Precision Education - A New Challenge for AI in Education — 105–108
Stephen J. H. Yang

Special Issue Articles

The Contribution of Dispositional Learning Analytics to Precision Education — 109–122
Dirk Tempelaar, Bart Rienties and Quan Nguyen

From Precision Education to Precision Medicine: Factors Affecting Medical Staff's Intention to Learn to Use AI Applications in Hospitals — 123–137
Hui-Chen Lin, Yun-Fang Tu, Gwo-Jen Hwang and Hsin Huang

A Result Confirmation-based Learning Behavior Analysis Framework for Exploring the Hidden Reasons behind Patterns and Strategies — 138–151
Fuzheng Zhao, Gwo-Jen Hwang and Chengjiu Yin

Toward Precision Education: Educational Data Mining and Learning Analytics for Identifying Students' Learning Patterns with Ebook Systems — 152–163
Christopher C. Y. Yang, Irene Y. L. Chen and Hiroaki Ogata

From Human Grading to Machine Grading: Automatic Diagnosis of e-Book Text Marking Skills in Precision Education — 164–175
Albert C. M. Yang, Irene Y. L. Chen, Brendan Flanagan and Hiroaki Ogata

An Analytical Approach for Detecting and Explaining the Learning Path Patterns of an Informal Learning Game — 176–190
Xuanqi Feng and Masanori Yamada

Predicting Students' Academic Performance by Their Online Learning Patterns in a Blended Course: To What Extent Is a Theory-driven Approach and a Data-driven Approach Consistent? — 191–204
Feifei Han and Robert A. Ellis

Twenty Years of Personalized Language Learning: Topic Modeling and Knowledge Mapping 205–222
Xieling Chen, Di Zou, Haoran Xie and Gary Cheng

Unfolding Students' Online Assignment Submission Behavioral Patterns using Temporal Learning Analytics 223–235
Mehmet Kokoç, Gökhan Akçapınar and Mohammad Nehal Hasnine

Determining Quality and Distribution of Ideas in Online Classroom Talk using Learning Analytics and Machine Learning 236–249
Alwyn Vwen Yen Lee

A Review of Using Machine Learning Approaches for Precision Education 250–266
Hui Luan and Chin-Chung Tsai

Analytics 2.0 for Precision Education: An Integrative Theoretical Framework of the Human and Machine Symbiotic Learning 267–279
Jiun-Yu Wu, Christopher C.Y. Yang, Chen-Hsuan Liao and Mei-Wen Nian

Using an Institutional Research Perspective to Predict Undergraduate Students' Career Decisions in the Practice of Precision Education 280–296
Tzu-Chi Yang, Yih-Lan Liu and Li-Chun Wang

Wen, Y., & Song, Y. (2021). Learning Analytics for Collaborative Language Learning in Classrooms: From the Holistic Perspective of Learning Analytics, Learning Design and Teacher Inquiry. *Educational Technology & Society*, *24* (1), 1-15.

Learning Analytics for Collaborative Language Learning in Classrooms: From the Holistic Perspective of Learning Analytics, Learning Design and Teacher Inquiry

Yun Wen[1*] and Yanjie Song[2]
[1]National Institute of Education (NIE), Nanyang Technological University, Singapore // [2]Education University of Hong Kong // yun.wen@nie.edu.sg // ysong@eduhk.hk
[*]Corresponding author

(Submitted June 30, 2020; Revised September 5, 2020; Accepted November 12, 2020)

ABSTRACT: Learning analytics (LA) has been increasingly using in teaching and learning. However, in the field of applied linguistics, the use of LA has only begun to touch the surface. There is a need for understanding how LA and learning design (LD) influence each other and provide useful information to language teachers in the context of specific courses or learning environments. In this study, a retrospective analysis was conducted to identify the factors influencing LA for collaborative language learning in classrooms, from a holistic perspective by integrating LA, LD, and teacher inquiry. The findings suggested that (1) LA focused on interactions can inform pedagogical refinement effectively when LD in language learning is premised on social constructivist theories; (2) LA supported teacher inquiry and LD on condition that the teacher holds innovation-oriented beliefs and the participatory culture between teachers and researchers. The study provided insights into the use of LA in collaborative language learning classrooms that focus on collecting and evaluating learners' interaction process beyond gleaning linguistic or behavioral facts. Professional development implications and future research are also addressed.

Keywords: Learning analytics, Learning design, Teacher inquiry, Second language learning, Collaborative language learning

1. Introduction

In recent decades, learning analytics (LA) as a nascent research field has aroused wide interest and attention in educational research and practice (Johnson et al., 2011; Rosé et al., 2016). It is an area of research involving collection, organization, analysis, and reporting of data about learners and their contexts in order to generate information and identify potential issues for prediction and pedagogical decision-making (Bienkowski, Feng, & Means, 2012; Ferguson, 2012; Reimann, 2016). Although the field is still in its infancy, its potential in the educational field has been widely acknowledged, including (1) explaining unexpected learning behaviors; (2) identifying successful learning patterns; (3) detecting misconceptions and misplaced effort; (4) introducing appropriate interventions; and (5) increasing users' awareness of their own actions and progress (Mangaroska & Giannakos, 2019; Papamitsiou & Economides, 2014). The process of using LA involves making sense of the presented information by taking contexts into account and taking actions accordingly (Siemens, 2013; Wise & Vytasek, 2017). Teachers can be engaged in LA by analyzing learning data for the purpose of informing, refining, and designing learning.

In the field of computer-assisted language learning, the evolution of techniques for analyzing big data is gaining attention, and LA has been used to evaluate linguistic data that can provide knowledge about learners' needs and areas of concern (Godwin-Jones, 2017), or analysis of students' behavioral patterns (e.g., Gelan et al., 2018). However, due to little agreement on what logged interaction data may be meaningful for understanding the complex learning process and enlightening language teachers, the use of LA has only begun to touch the surface in the field of applied linguistics (Godwin-Jones, 2017; Link & Li, 2015). Meanwhile, Greeno (1998) and others have argued that any analysis of learning will be incomplete if it does not conceptualize learning as a sociocultural practice. Reimann (2016) also further stated that such an understanding of learning practices is necessary for not only theoretical but also pedagogical purposes. Hence, there is a need for understanding how LA can provide useful information to students, teachers, and designers in the context of specific courses or learning environments. On this basis, a holistic perspective by integrating LA, learning design (LD), and teacher inquiry, is particularly important to investigate the potential of LA in the computer-supported language classroom, where language learning is distributed, and creating and sharing content by learners are enabled by web 2.0.

Therefore, this paper shares a case of computer-supported collaborative Chinese second language (L2) learning in Singapore. McKenney and Mor's (2015) integrated framework of LA, LD, and teacher inquiry is adopted to articulate how they are mutually supportive and how LA could be beneficial to collaborative language learning in classrooms. The study seeks to provide insights into the integration of LA in the computer-supported collaborative language learning classroom that focuses on collecting and evaluating learners' interaction process. Implications for the development of future LA tools and professional development for language teachers are also explored.

2. Literature review

2.1. LA in language learning

Big data have played a significant role in the field of computer-assisted language learning and second language acquisition (Godwin-Jones, 2017). Advances in natural language processing have enabled rich tagging and annotation of corpus data, and there has been growing interest in learner corpora and student-created corpora, collections of written or spoken language by learners, to highlight frequent or common errors in students' lexis or syntax (Godwin-Jones, 2017). With advances in LA, educators' decision-making process is informed with actionable information for supporting students' language learning. Li and colleagues (2018) used LA to examine learners' self-regulated behaviors based on logged data from 2454 freshman university students, and they identified typical learning patterns in an online learning environment which inform educators of pedagogical decision-making. Based on a computer adaptive testing platform, Aristizábal (2018) showed how an American international school in Vietnam had been using data and LA to understand students' learning from different forms of assessment and how they used these findings to improve the reading skill of English as a foreign language students. The study pointed out that it was teachers' pedagogical and epistemological backgrounds that made LA significant and meaningful, which contributed to teachers' reflection on their teaching practices and gaining insights into pedagogical decision-making to improve student English reading.

Some studies have shown the effectiveness of LA in computer-assisted vocabulary learning, and in these studies, language learning was conceptualized as the sociocultural practice where contexts and interaction were emphasized (Ellis, 2008). For example, Mouri et al. (2018) reported a study on evaluating the effectiveness of the learning analytics tool VASCORLL 2.0 in connecting students' vocabulary learning acquired via eBook to that learned from real-life in higher education. VASCORLL 2.0 helped to automatically visualize and analyze all learning logs accumulated in an e-Book system and real-life learning. For instance, if a learner learned and saved a new word using the e-Book, the learner's node and its links with the relevant node of word or learning material in the learning structure would be visualized. The findings indicated that the tool is effective in providing learners' more opportunities in vocabulary learning. In addition, Hsiao, Lan, Kao, and Li (2017) developed a visualization analytic method to understand the impact of various learning strategies on college students' Chinese vocabulary in a virtual world. The research findings show that the visualization analytics method could help teachers visualise students' different learning strategies of vocabulary acquisition.

There is also a trend to apply LA to students' collaborative writing. Hu (2017) proposed a LA tool embedded in Wikiglass for automatically recognising, aggregating, and visualising levels of thinking orders in secondary student collaborative writing in inquiry-based learning in order to assist teachers in identifying individual and group work and selecting example sentences as teaching materials. Text categorisation models were constructed and evaluated using machine learning and natural language processing techniques, to make an attempt to identify at-risk individuals and groups, refine assessment rubrics, select example sentences as teaching materials, and facilitate students' self-regulated learning. However, as Yim and Warschauer (2017) stated in their paper about text mining used in web-based L2 collaborative writing, the use of data and text mining for understanding writing processes in language learning contexts was largely underexplored.

In sum, the majority of studies on LA in language learning have been conducted in higher education and concentrated on exploring the role of LA in students' self-regulated learning. In addition to evaluating linguistic data, the potential of LA used in understanding the complex interaction process in sociocultural views of language learning, and its significance of pedagogical implications have been increasingly acknowledged. There is scarce literature reporting how to analysing and visualising interaction data in collaborative language learning in classrooms, though the importance of quantifiable information about collaborative language learning has been highlighted (Yim & Warschauer, 2017).

2.2. Interplay of LA, LD and teacher inquiry

2.2.1. Interplay of LA and LD

The relationship between LA and LD can be traced back to collecting qualitative and quantitative data to inform educators of revision and fine-tuning of the instructional system under development before the terms LA and LD were coined (Persico & Pozzi, 2015). LD is another field associated with technology-enhanced learning. It can be seen as a form of documentation of pedagogical intent that can provide the context to make sense of diverse analytic data (Lockyer, Heathcote, & Dawson, 2013). Today's classrooms are turning to be technology-rich and data-rich environments, where teachers need to make effective use of technology. Yet during the process of integrating technology with teaching in classrooms, teachers may generate a series of "wonderings" or "burning questions" from practice. To keep the pace of the fast development of technological tools and their use in teaching and learning, today's teachers, more than ever before, need to change the traditional classroom teaching mode, and break their isolation to develop more solid, extensive, and dynamic design competence (Conole & Culver, 2010; Persico & Pozzi, 2015). Hence, a collaborative inquiry endeavor by teachers is underscored by research in LD (Laurillard, 2013; Mor & Mogilevsky, 2013; Persico & Pozzi, 2015).

As argued by Lockyer et al. (2013), LD is based on its reusability and adaptivity across educational contexts, in which excellent teaching practice can be captured, interpreted, and readapted. LD could support teachers as designers of a technology-enhanced learning environment and enhance their professional expertise (Voogt et al., 2011). Both research and the practice of LD have evolved with mainly two purposes: to provide models or a framework to promote teaching quality with shared knowledge among educators (Lockyer et al., 2013) and to focus on the integration of technology into teaching and learning for a semantic structure for analytic (Mor et al., 2015). Several studies also have tried to improve the LD experience by utilizing LA and providing a contextual framework to help teachers understand the information that LA provides (Bakharia et al., 2016; Lockyer et al., 2013; Sedrakyan et al., 2018) so as to further facilitate pedagogical actions.

2.2.2. Mediating role of teacher inquiry

The mediating role of teacher inquiry when connecting LA with LD has also been acknowledged (Alhadad & Thompson, 2017). Teacher inquiry can be seen as a set of research practices by which teachers generate questions, collect data, as well as examine their practice and effects on students' learning for enhancing their professional knowledge and improving teaching practice in the classroom (Clarke & Erickson, 2003). Unleashing the potential of teacher inquiry for improving teachers' practice and professional developments supported by LA and LD is significant in technology-rich environments. The inquiry process is characterised as "purposeful observation that involves deliberate planning to anticipate what, when, and how methods are to be observed" (Rich & Hannafin, 2008, p. 1427) and one of the goals of teacher inquiry is to develop educators' ability for critiquing and improving their own teaching. Recently, some research studies have focused on the opportunities provided by technology, such as LA to support holistic and data-informed LD and teacher inquiry. Mor et al. (2015) stated that teacher inquiry should be a method for data-driven teaching reflection for the benefit of student learning, which is aligned with Dawson's opinion (2006) that teacher inquiry could be recognized as the process of helping teachers to learn to become effective technology-users. LA is able to make teacher inquiry and LD more scientific and reliable. However, the "intuitive and tacit practices of LA, LD and teacher inquiry are far from the structured and technology-enhanced vision of these three domains common in the academic literature." (McKenney & Mor, 2015, p. 265).

2.2.3. Investigating LA in language learning from a holistic perspective

To sum up, the fields of LA, LD, and teacher inquiry do not occur in isolation (McKenney & Mor, 2015), but need to be studied synergistically. As stated by Mor et al. (2015), LA could be used to improve LD and support teacher inquiry. LD provides a semantic structure for LA, whereas teacher inquiry could utilize the information provided by LA (McKenney & Mor, 2015), as well as raise reflective questions for LA and LD to follow up. Regarding the integration of LA, LD, and teacher inquiry, some studies have focused on technology development (Haya et al., 2015; McKenney & Mor, 2015; Mouri, Uosaki, & Ogata, 2018), and adopting top-down design principles (Bakharia et al., 2016; Ifenthaler et al., 2018). Other studies have proposed an integrated conceptual framework (McKenney & Mor, 2015; Mor et al., 2015; Persico & Pozzi, 2015) to illustrate the relationships among them. However, studies that have been conducted to investigate the synergetic relationship among them empirically are rarely found, particularly in the context of language learning. The holistic perspective by them can help to understand LA for collaborative language learning in classrooms more comprehensively and meaningfully.

3. Research questions

From the holistic perspective of LA, LD, and teacher inquiry, this study aims to understand how LA can be beneficial to collaborative language teaching and learning in the classroom environment, taking a case of computer-supported collaborative L2 writing as an example. Cobb et al. (2003) argue that retrospective analysis "provides a trustworthy account of the process whereby a series of events—each of which is local and contingent—can be seen as part of an emergent and potentially reproducible pattern" (p. 12). Retrospective analysis attempts to generate a coherent framework that counts for situated effects, thus making it possible to anticipate outcomes in future designs (Cobb et al., 2003). Therefore, in this paper, the factors influencing LA for language teachers in the computer-supported collaborative learning are systematically identified via retrospective analysis of data collected in the GroupScribbble (GS)-supported collaborative L2 learning study. The retrospective analysis is guided by the two research questions below.

(1) How LA focusing on interactions and LD were mutually beneficial in computer-supported language classrooms?
(2) What factors influenced and shaped teacher inquiry when connecting LA with LD?

4. The GS study

The case of GS-supported collaborative writing with a learning analytics tool in a Chinese L2 class was deemed appropriate for this retrospective analysis. First, the case synergistically combined the three key elements, LA, LD, and teacher inquiry. Second, the effectiveness of using GS in L2 classrooms to enable productive collaborative learning, and the essential role of teachers during the process had been evidenced (Wen, 2019). Data from the GS intervention were re-coded to understand how its LA tool interacted with the LD and mediated by teacher inquiry.

4.1. Participants and the computer-supported classroom enabled by GS

The case was about a Chinese language teacher under the pseudonym "Mdm Y" who used an online representational tool—GS in her secondary 3 (Grade 9) Chinese writing class in a neighbourhood school. The teacher had more than 10 years of teaching experience. The class was a normal Chinese class with 29 local students (aged between 15 to 16 years). In most secondary schools in Singapore, students are channeled into a higher Chinese or normal Chinese class based on their language proficiency. Compared with the higher Chinese class, the normal Chinese class requires a comparatively lower level of linguistic proficiency and cultural knowledge.

GS is a generic online representational tool for supporting brainstorming and knowledge construction, like Lino or Padlet that has been widely used in today's classrooms. In a typical GS class, students were assigned to small groups with 3 to 5 members, sitting together and using a personal computer. In this way, students in the same group were able to do both GS-based and face-to-face interactions. The class of students was classified into 7 groups homogenously to complete pre-writing activities using GS in the classroom environment. They were grouped homogeneously according to their scores of the Chinese language examination at the end of Secondary 2 when it was before intervention. The total score of the examination was 100, and the average score of the class was 57.03 (*SD* = 9.13). As students' language proficiency might influence their reactions to different pedagogical approaches (Wen, 2019), we sought to purposefully investigate a high-ability group and a low-ability group to maximize the variation of interaction patterns. Based on that, a group with lower language proficiency (Group 1) and a group with higher language proficiency (Group 6) were randomly selected as our target groups. They were Group 1 (*Mean* = 51.14, *SD* = 3.07) and Group 6 (*Mean* = 62.47, *SD* = 3.88).

As shown in Figure 1, the GS user interface presents each user with a two-paned window. Its lower pane is made up of the user's private board, whereas the upper pane is the public board. The private board is provided with virtual pads of fresh scribble sheets on which the user can draw or type. Students can share the scribbles sheets by dragging them from the private space to the public space. The most essential feature of the GS is the combination of the private board, where students can work individually, and group boards, where students can post the work, view others' work and take items back to the private board for further elaboration. A student can select any group board by clicking on the board number at the top right corner, and browse all other groups' postings on the public space. Hence, the tool may make intra- and inter-group interactions more convenient. In this way, students have an opportunity to exchange their ideas and provide comments for one another without physical movement in classroom environments.

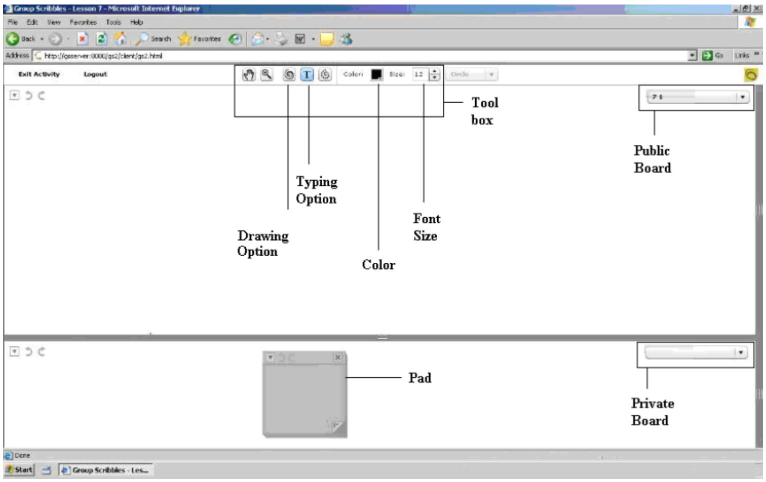


Figure 1. The user interface of GS with a two-paned window

After technical training for both the teacher and all the students, we also conducted a series of professional development sessions for the teacher to ensure that she had a good understanding of GS-based language learning design. In the professional development session, the rapid collaborative knowledge improvement (RCKI) concept and its related design principles (e.g., spontaneous participation, multimodal expression, or idea diversity) were introduced. RCKI focuses on democratic knowledge sharing and continuous individual and group knowledge enhancement (Wen, Looi & Chen, 2011; Wen, 2019). It is proposed to address the constraints faced by classroom teachers when they are designing and implementing knowledge construction and improvement practices within the short duration of a classroom lesson.

4.2. GS-based collaborative writing activity design

Collaborative L2 writing is a recursive, bottom-up process that requires participants to collaboratively contribute words/phrases or ideas, and eventually compose their compositions in individuals or groups (Oxford, 1997). Although brainstorming as the fundamental pre-writing activity has been emphasized in various models of the writing process, there is a paradox between L2 writers' ideas expressions and their limited target language proficiency. Some studies (Scott, 1996; Stapa & Majid, 2009) have shown that, for L2 writers, the low proficiency of the target language often requires them to focus primarily on vocabulary and grammar, and hence hampers idea generation and expression. Nonetheless, there are also studies (e.g., Wong et al., 2009) suggesting that the collaborative writing process should start with contributing vocabulary. In practical collaborative L2 writing activities, teachers usually encourage students to write down their ideas or vocabulary as much as possible.

Therefore, both "idea first" and "vocabulary first" strategies were used in designing GS-based collaborative writing activities. In the "idea first" activity, students were encouraged to contribute ideas relevant to the topic using whichever representational forms they are comfortable with and competent in. For example, they could share their ideas in English or drawing pictures. In the "vocabulary first" activity, students were encouraged to contribute Chinese words, phrases, or idioms they could think about with the given topic (more details can be found in the section of results).

The "idea first" activity was designed for Lesson 1 and Lesson 2 with the writing topic "why do Singaporeans feel discontented?" In Lesson 1, students were required to start writing by brainstorming ideas on the GS platform. After that, in Lesson 2, students continued to organize the ideas they shared and select a few good ones to further refine and expand to paragraphs via GS. Taking the screenshot of group 1 in Lesson 1 as an example (Figure 2), after brainstorming in term of the given topic, the group of students classified their ideas into three main categories, "claims," "reasons" and "suggestions/solutions" in term of the four elements provided by the teacher. One week later, the same group in Lesson 2, further refined and improved their group artifacts. As shown in Figure 3, three main reasons proposed in the prior lesson were selected and further elaborated and their corresponding solutions were provided and improved.

Figure 2. GS screenshot from Group 1 at the end of Lesson 1

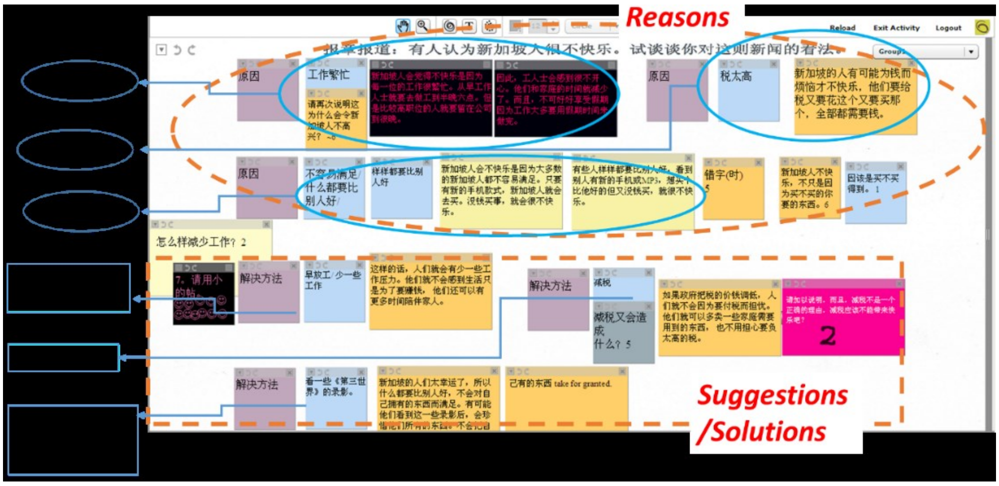


Figure 3. GS screenshot from Group 1 at the end of Lesson 2

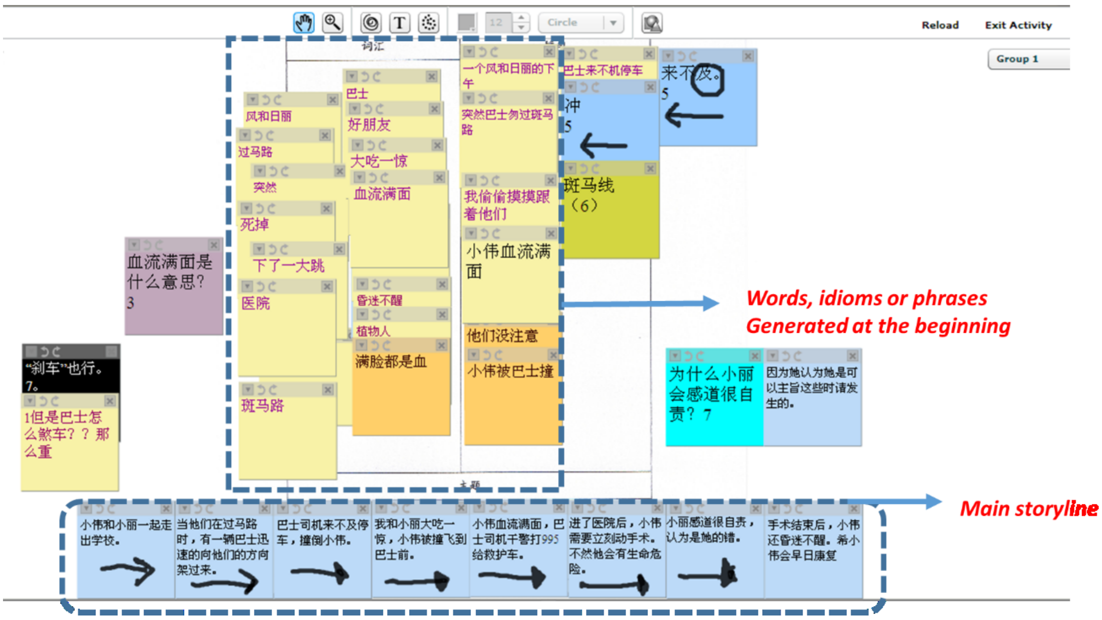


Figure 4. GS screenshot from Group 1 at the end of Lesson 3

The “vocabulary first” activity was designed for Lesson 3 and Lesson 4, in which the task was to make a story based on the given scenario that “when you see your classmates whispering at the school gate, what will you think of?” In Lesson 3, students were asked to start writing by generating vocabulary related to the new topic (Figure 4). In Lesson 4, they were asked to generate main paragraphs with the words/phrases collected in Lesson 3. In the “words/phrases” activity, students were encouraged to contribute suitable Chinese words, phrases, or idioms that could be adopted directly in the final writing. In the process of interacting with others, students could enlarge their vocabulary and equip themselves with better understandings of the collected vocabulary.

4.3. LA module for GS

An analytical module for GS was designed for teachers and researchers to capture and analyse students’ GS-based actions, such as posting, editing, or visiting others’ postings. It could be used both during and after class. With the analytical tool, the actions of each student could be captured and recorded automatically by the data logging mechanism. A large amount of logged data was converted to indicators and visualized to show whether students actively engaged in the learning activity. Underlying the pedagogical concept of RCKI, seven types of action were included in the analytical tool (see Table 1). For instance, the number of “post to board” can indicate whether students participate in idea sharing actively. Whether they have awareness of improving their own ideas (through the number of “edit own post on board”) or modifying answers for others (through the number of “edit another post on board”) can be reflected as well. The unit of analysis can be an individual student or a group of students.

Table 1. Action indictors based on the logged data.

No.	Action	Description
1	Post to board	Create a scribble note and post (move) to the public board
2	Edit own post on board	Edit the content of the scribble note that is posted on the public board. The scribble note is posted by the same user.
3	Edit another post on board	Edit the content of the scribble note that is posted on the public board. The scribble note is posted by a different user.
4	Repost own post to board	Move the scribble note that is posted on the public board to the private board, and then put it back on the public board. The scribble note is posted by the same user.
5	Repost another post to board	Move the scribble note that is posted on the public board to the private board, and then put it back on the public board. The scribble note is posted by different users.
6	Delete own post from board	Delete the scribble note from the public board. The scribble note is posted by the same user.
7	Delete another post from board	Delete the scribble note from the public board. The scribble note is posted by a different user.

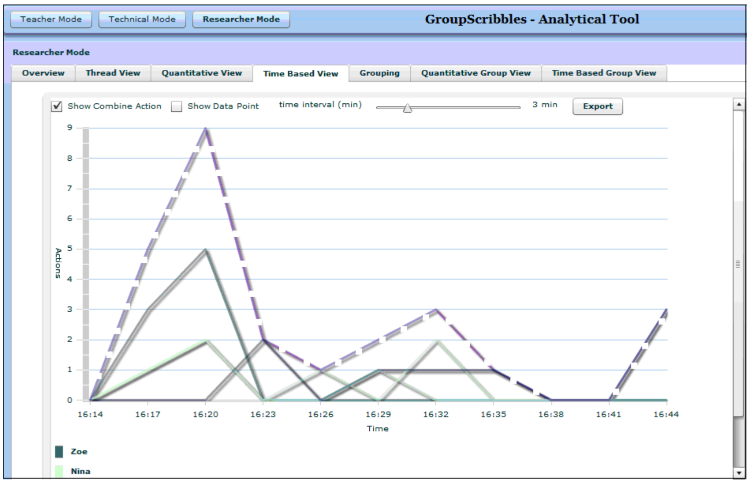


Figure 5. Time-based actions of a group of students

Figure 5 is a screenshot of the analytic tool’s interface. It displays time-based actions of a group of students, via “Time Based View.” In the graph, the X-axis refers to time and the Y-axis refers to the number of actions. The solid lines in different colors represent different student actions and the dashed line reflects the time-based actions of the whole group. It can be read from the graph that the GS-based activity started from 16:14 and lasted for 30 minutes until 16:44. After the point of 16:23, the number of group actions dropped to a low level. The group actively participated in the first 9 minutes of the activity.

4.4. Data sources

In order to reduce the risk of bias and enhance the trustworthiness of our findings, we re-coded the existing data from multiple sources and collection methods (Lincoln & Guba, 1985). More specifically, there were five types of data: (1) lesson plans; (2) field notes from each post-lesson discussion; (3) teacher's interviews; (4) screenshots of LA findings; (5) screenshots of student-generated artefacts. We then triangulated these multiple sources of data.

(1) ***Lesson plans***: Before every GS lesson, Mdm Y designed a lesson plan and shared it with the research team, based on the RCKI concept and its related design principles. The research team provided feedback and suggestions to her before each lesson. A total of 8 lesson plans about GS-based collaborative writing were co-design.

(2) ***Field notes from each post-lesson discussion:*** We immediately conducted a post-lesson discussion with Mdm Y after each GS lesson. The purpose was to encourage the teacher to reflect on her lesson design and enactment. During the process, we shared class observation notes and LA results drawn from the analytical tool with her. The first author of the research team led these discussions, and all members of the research team (usually 2-3 researchers) kept written field notes of the discussion.

(3) ***Teacher interviews:*** We had two formal interviews with the teacher before and after our collaborative respectively. Both interviews were semi-structured and conducted by the first author and lasted about 50 mins. There were some of the same questions in the pre-and post-interview, in order to track the teacher's belief, her understandings of pedagogical design, and the use of LA. The interviews were audio-recorded and verbatim transcribed.

(4) ***Screenshots of LA visualization*:** The analytical module for GS was developed and used after the first semester of the intervention, so it was used to analyze the four GS lessons in the second semester. Some representative screenshots of LA findings of the four lessons were saved and shared with the teacher.

(5) ***Screenshots of student-generated artefacts:*** After every GS lesson, the research team helped Mdm Y to do a screenshot of each group's artifact. Then, Mdm Y printed out these screenshots and shared them with students as supporting materials for subsequent individual writing. In this paper, we did not investigate the process and effect of the learning activities via these student-generated artefacts, because that was not the focus of this paper. We mainly used these screenshots to explain the activity designs and to support the findings drawn from the LA tool.

4.5. Data analysis

For the present study, summarized data from the original set were re-revisited in light of the framework of LA, LD, and teacher inquiry. Narrative analysis was used to interpret the connection between LD and LA, in terms of lesson plans, field notes, screenshots of LA visualizations and student-generated artifacts in a storied form (Webster, & Mertova, 2007). In order to identify the factors that influence and shape teacher inquiry when connecting LA and LD, thematic analysis was adopted to identify and analyze patterns of meaning from field notes and teacher interviews (Clarke & Braun, 2017). The first author identified the recurring themes in the data set that reflect the teacher's beliefs, her understanding of pedagogical design, and the use of LA. The second author helped to check and add inputs for revisions. Data sources and analytic approaches are presented in Table 2.

Table 2. Data sources and analytical approaches

Data source	Analytic approach	Q1	Q2
(1) Lesson plans	Narrative analysis	x	x
(2) Field notes	Thematic analysis	x	x
(3) Teacher interviews	Thematic analysis		x
(4) Screenshots of LA visualization	Narrative analysis	x	
(5) Screenshots of student-generated artifacts	Narrative analysis	x	

5. Results

5.1. LD can guide the design and interpretation of LA on interaction, which in turn inform pedagogical decision-making

Drawn from the GS analytical tool, Figures 6 and 7 represented student's participation at the class level in Lesson 1 ("ideas first" activity) and Lesson 3 ("vocabulary first" activity) respectively. The horizontal axis of the figure represents the time of the class period with an interval of 6 minutes. The vertical axis is the aggregate count of all students' GS-based actions including posting, editing, reposting as well as deleting. Some logged data such as creating a scribble pad without content, undoing or moving were excluded. In terms of the total number of GS-based actions, there was an obvious difference between the "idea first" lesson (N = 10,030) and the "vocabulary first" lesson (N = 9,851). According to the RCKI principles, at the beginning stage of the GS-based activity, students should be encouraged to post their ideas the more the better. The analytics visualization indicated that students were more actively engaged in the "ideas first" activity than in the "words first" activity in the first 6 minutes of "brainstorming."

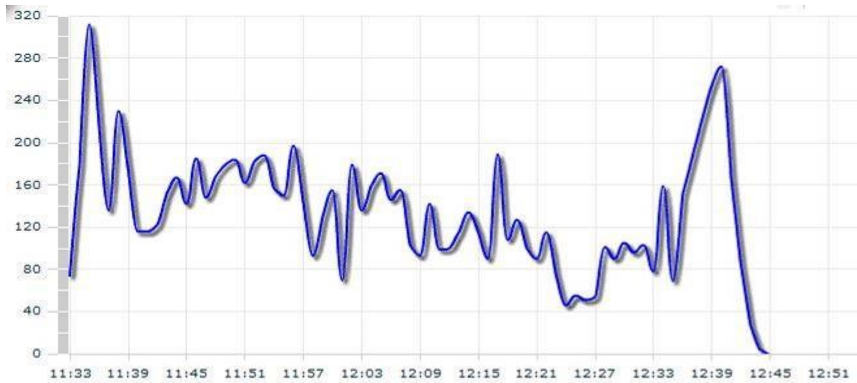


Figure 6. GS-based actions in the "ideas first" activity in Lesson 1 at the class level

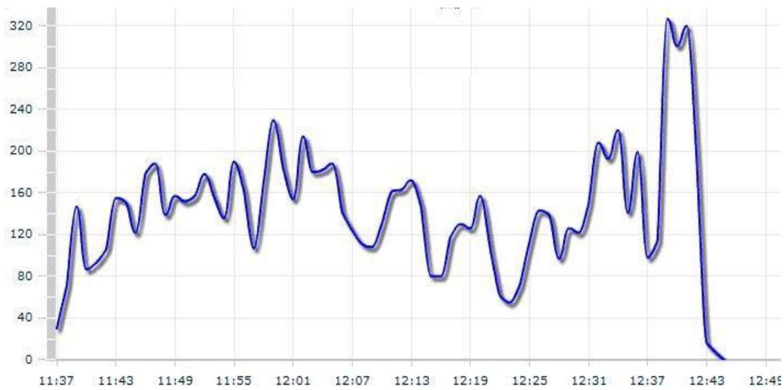
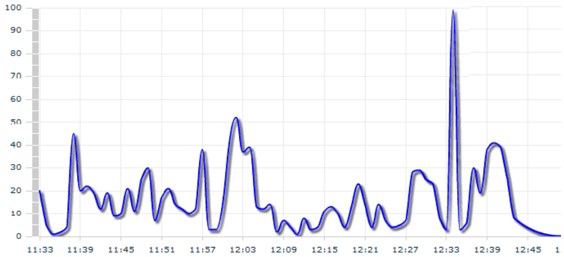


Figure 7. GS-based actions in the "vocabulary first" activity in Lesson 3 at the class level

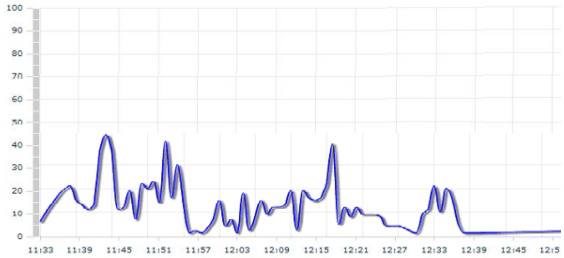
The GS analytical tool could help teachers to visualize students' participation at the group and individual level as well. Here we showed the findings at the group level to interpret them based on the pedagogical designs. Figure 8 compares the participation levels of the two target groups (Group 1 and Group 6) in the "ideas first" and "vocabulary first" activities. As shown by the LA results, different pedagogical approaches of "ideas first" and "vocabulary first" had an impact on the participation level of groups with different language proficiency. Group 1 with lower language proficiency (an average of 20 actions per 6 minutes) was even more actively engaged in the "ideas first" activity than Group 6 with higher language proficiency (an average of 17 actions per 6 minutes). Nevertheless, the difference in participation between the two groups was not prominent in the "Vocabulary first" activity (both averaged18 actions per 6 minutes).

In addition, no matter in Lesson 1 or Lesson 3, a massive jump of actions could be observed in Group 1, but it was not obvious in Group 6. The two peaks occurred at the time that the teacher asked the students to further refine their own postings, after visiting other groups' artifacts and providing comments for others. The LA findings illustrated that compared with Group 6, Group 1 with lower language proficiency revised their work more intensively. Yet it was indeed uncommon that nearly 100 actions occurred in such a short time in a group.

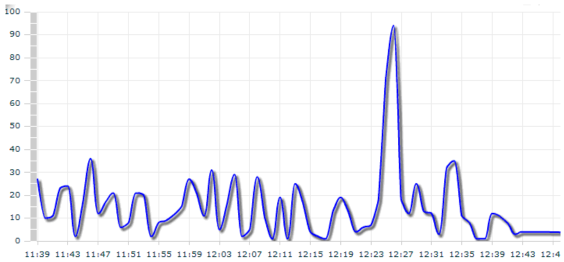
This might also have something to do with the group's habit of using GS. They might drag all the postings back to the private board and then repost them to the public board for re-organization. Nevertheless, it was a good sign that the group was working actively as a group to improve their group's artifacts. In this sense, it might be further interpreted that further scaffoldings at the social plane should be provided to the students in the high ability group, to encourage them to participate in the collaborative knowledge improvement more actively and improve their knowledge consistently. On the other hand, further scaffoldings at the cognitive plane were more essential to the students in the low ability group to make sure their more sustained knowledge improvement.



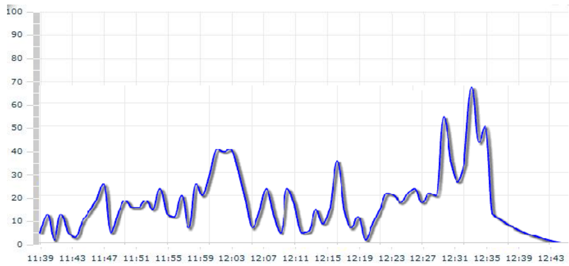
(a) Actions of Group 1 in Lesson 1



(b) Actions of Group 6 in Lesson 1



(c) Actions of Group 1 in Lesson 3



(d) Actions of the Group 6 in Lesson 3

Figure 8. GS-based actions in the "ideas first" (a and b) and "vocabulary first" activities (c and d) at the group level

The lower-ability students typically had very low motivation to participate in Chinese writing. After the researchers shared with Mdm Y the LA results, she said "Initially, I did not expect the lower ability group would complete the task well." Moreover, to the surprise of the teacher and the researchers, the GS-based group artifacts generated by the lower-ability group were even better than those generated by the higher-ability group in the "idea first" activity, in terms of the richness of and logic of ideas (Group 1's final artifact was shown in Figure 4, and Group 6's artifact was shown in Figure 9). Mdm Y said with the feedback, she would be more confident in providing students more time to reflect, involve, and revise their artefacts.

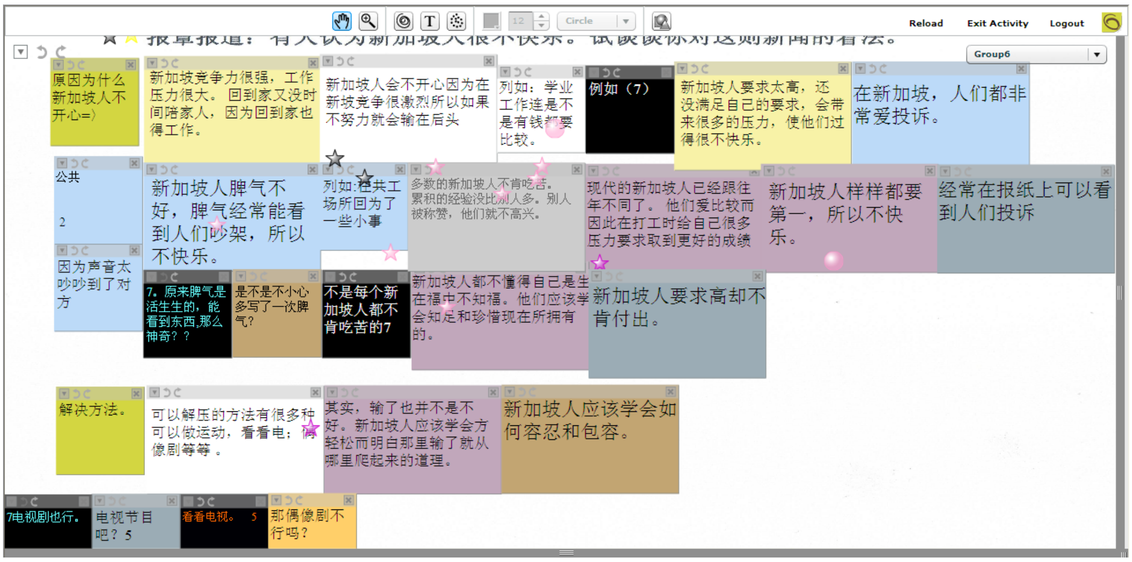


Figure 9. GS screenshot from Group 6 at the end of Lesson 2

The case demonstrated the synergistic relationship between LD on interactions and LA. Although the GS LA module was far from an ideal analytic tool, its findings concentrating on students' social participation at multiple levels (class and group levels) could help teachers to monitor students' learning process and make pedagogical decisions. For instance, what length of time should provide for students' online brainstorming; more supports should be provided to the lower ability group at the stage of refined their group artifacts; as well as the "idea first" approach was more appropriate to L2 learners than the "vocabulary-first" approach in collaborative writing.

5.2. Factors influenced and shaped teacher inquiry when connecting LA with LD

5.2.1. Teacher beliefs about and capacity for pedagogical innovation

The interview data demonstrated that Mdm Y held the belief of adopting an innovative approach in her Chinese language teaching. She had motivation and aspiration to change the traditional teaching approach, and she was open to new teaching approaches. As she shared with the research team, even though sometimes the outcomes of innovative approaches were not very positive, she still believed that the diversity of teaching approach helped to improve students' language learning interests. Yet she also mentioned that she was conservative on using technology before the GS lesson. As an experienced teacher, it was not difficult for her to design the first GS lesson. After our first professional development session, she showed confidence in the activity design and shared her lesson plan with us in a short time. The first lesson in the first semester was about describing people and personal qualities, focusing more on vocabulary. However, soon after, she began to realize that in this kind of design, students were difficult to be involved in consistent deep interactions after sharing vocabulary. That was why the pedagogical question was proposed, and the comparison between the approaches of "idea first" and "vocabulary first" were conducted.

When Mdm Y was asked what the biggest challenge she faced in designing GS lessons in the post-interview. She said "the biggest challenge is to be able to reflect the principles of using GS. That is the first point. The other point is how I select a proper topic… I need time to… slowly, I mean, to just come up with the right design for the lesson plan." It demonstrated that the teacher gradually had a better understanding of the design principles of rapid collaborative knowledge improvement, and she began to realize that there was a big difference between the learning activity design emphasized collaborative knowledge construction and the learning activity design mainly concentrated on the transmission of content knowledge. The teacher was eager to learn as she was keen on making pedagogical innovations.

5.2.2. Participatory culture between teachers and researchers

Mdm Y attributed the success of her lesson designs to the collaboration with the research team. She said in the post-interview "I spent a lot of time thinking about what if they (students) were not collaborating. So it is important that I work with you (the first author). Basically, you are half the brains behind this. In the course of (planning) lessons. We are indeed collaborating on this."

Meanwhile, although Mdm Y felt the LA findings we shared with her in post-lesson discussions were interesting and helped her to be more confident to make pedagogical decisions, she also mentioned that she did not have much time to check LA results in the class. It was not easy for her to interpret the results and she was overwhelmed by answering students' answers and observing their posting on GS boards. According to her feedback, it was more straightforward for her to monitor students' learning via viewing students' artifacts shared in the public boards. However, she also mentioned that the LA tool helped to provide her with a holistic picture of student participation. Without the LA tool, it was hard to capture these pieces of information as students were keeping on posting on the GS board during the activities. With the researchers' elicitation and facilitation in how to interpret the LA results based on rapid collaborative knowledge improvement design principles during the post-lesson discussions, the teacher was able to understand how to take advantage of LA on interactions to monitor learning processes and even predict collaborative learning effectiveness.

6. Discussion

6.1. LA on interactions and LD were mutually beneficial in language learning

The GS-based study showed that LA on cognitive interactions could make group participation at multiple levels in the collaborative learning process visualized. The results of LA informed LD regarding the pedagogical strategy ("idea first" strategy in this case) that could better be used for collaborative language learning. In the classroom environment, informed by LD, LA on interactions at the group level might be more effectively help teachers monitor learning processes and make pedagogical decisions.

The majority of existing studies on LA in LD has focused on distilling common linguistic mistakes that students have made (Godwin-Jones, 2017), identifying student behavioral patterns (Gelan et al., 2018), examining student self-regulated learning (e.g., Li et al., 2018), or investigating LA's impact on student academic performance (e.g., Mouri et al., 2018), especially in higher education. Although Hu's (2017) study adopted LA in enhancing learners' collaborative writing, it focused on assessment and identifying at-risk students instead of analyzing social interactions at multiple levels. This present study was unique in that it demonstrated how LA on interactions and LD were mutually beneficial in language learning in classrooms.

In addition, it is also noted that in this study, the synergy between LA and LD lies in the visualization of interactions in the social constructivist theory-led design of collaborative learning activities, which has rarely been found in current LA studies on collaborative writing LD. The example of LA on interaction at the pre-writing stage underpinned by rapid collaborative knowledge improvement design principles shows that LA can better inform pedagogical decision-making when collaborative writing LD is grounded with sound theories. According to Yim and Warschauer (2017), this is the research area that still needs to be further explored.

6.2. Teacher inquiry mediating the connection between LA and LD

The GS-based case study provided an example to explain how LA supported teacher inquiry to make evidence-based pedagogical decisions. The L2 teacher defined a burning question ("Idea first" or "vocabulary first," which approach is more effective in GS-based collaborative L2 writing) that emerged from her practice of using a GS for improving students' collaborative writing. Indeed, the findings towards the first research question of this study suggested that the LA results were instrumental in making pedagogical decisions. Yet the teacher played a mediating role, and the process of teacher inquiry depended on factors regarding teacher beliefs, a collaboration between researchers and the teacher, and the usefulness of the visualization results.

As the teacher held beliefs of adopting the innovative pedagogy in her teaching, she was keen on collaborating with the researchers and being involved in LD. However, she encountered some difficulties in designing and enacting GS-based activities underlying the rapid collaborative knowledge improvement, and her self-use and interpretation of LA results were quite limited. There might be two possible reasons. One was that this technology-enhanced collaborative writing approach was still new to the teacher. Even with the collaboration between the research team, the teacher still needed time to internalize her understanding of learning activity designs. The cognitive load on the teacher is high during enactment of the computer-supported collaborative learning activities (Prieto et al., 2014). The other reason was because of the frustration in understanding the visualized LA results. Asynchronous analytics interpreted together with the research team, helped the teacher to compare group performances, and make a pedagogical decision. LA, both in asynchronous and synchronous forms can provide opportunities for insights into the language learning process (Reinders, 2018). Ideally, in this case, the teacher would be able to use synchronous analytics to provide more targeted supports to groups in the classroom environment. LA tool introduced in the study still has room for improvement both in functionality and visualization. Future research on teacher inquiry and LA needs to be conducted based on more powerful and user-friendly LA systems and pay attention to both synchronous and asynchronous forms of LA.

Despite the recognition that teachers should be equipped with data for teacher inquiry, there is a dearth of research on how to improve the capacity of teachers to use data (Mandinach & Jimerson, 2016). Existing studies focus on investigating teachers' beliefs about data use (e.g., Datnow & Hubbard, 2016; Reeves & Chiang, 2019). The findings of this study suggest that teachers' beliefs about the innovative pedagogical approach may affect their beliefs about data use. On the other hand, existing studies show that innovation-oriented teacher beliefs and participatory culture can influence teacher inquiry and LD (e.g., Song & Looi, 2012). This study demonstrates that the participatory culture between teachers and researchers should be highlighted across the processes of LD, transforming data into information, and transforming information into a decision.

6.3. Limitation of this study

There is no denying that this study has some limitations. First, as we discussed, the LA tool introduced in the study is far from satisfactory in visualizing interactions for RCKI. The retrospective analysis of the study aimed to provide some insights into the design of LA tool to enable a fine-grained analysis of the collaborative writing process based on social constructivist theories. Secondly, investigating the interplay of LA, LD, and teacher inquiry should be a longitudinal and cyclic process. The case shared in this paper did not go further to collect the data from redesigning and implementing refined learning activities. Further research needs to focus on empirical studies to examine the synergistic relationships among LA, LD, and teacher inquiry longitudinally and at a larger scale to enhance L2 classroom teaching and learning.

7. Conclusion and implications

This study investigated the use of LA in language learning classrooms via exploring the synergistic relationships among LA, LD, and teacher inquiry in a case of collaborative L2 writing. The findings show that (1) LA focused on interactions could inform pedagogical refinement effectively when LD in language learning was premised on social constructivist theories; and (2) LA supported teacher inquiry and LD on condition that the teacher held innovation-oriented beliefs and was keen on collaboration with researchers and professional development.

The implications of this study are twofold. Firstly, as LA is concerned with how to provide teachers with relevant, understandable, and actionable information (Rodríguez-Triana et al., 2018), the study suggests that the indicators and visualization of LA may need to be more straightforward to reflect the collaborative writing process underpinned by sound theories, which, in turn, can inform teacher inquiry and LD. Secondly, teachers' beliefs about the innovative pedagogical approach and teacher-researcher collaboration are critical for teacher inquiry in innovative language LD and practices, which can help the teacher make sense of LA for pedagogical refinement to improve interactions in collaborative language learning. Teachers' professional development on innovative pedagogies should take precedence over their professional development of data literacy.

Acknowledgement

This study is funded by Education Research Funding Programme, National Institute of Education (NIE), Nanyang Technological University, Singapore, with project no. OER 02/20 WY. The data for retrospective analysis is based on the work supported by the National Research Foundation (Singapore) under Grant NRF2007-IDM003-MOE-001. We are grateful to all the researchers, teachers, and students participated in the project.

References

Alhadad, S. S. J. & Thompson, K. (2017). Understanding the mediating role of teacher inquiry when connecting learning analytics with design for learning. *Interaction, Design, & Architecture, 33*, 54–74.

Aristizábal, J. A. (2018). Using learning analytics to improve students' reading skills: A Case study in an American international school with EAL students. *GiST Education and Learning Research Journal*, *17*(17), 193–214.

Bakharia, A., Corrin, L., De Barba, P., Kennedy, G., Gaševíc, D., Mulder, R., Williams, D., Dawson, S., Lockyer, L. (2016). A Conceptual framework linking learning design with learning analytics. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 329–338). doi:10.1145/2883851.2883944

Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An Issue brief.* Washington, DC: SRI International.

Clarke, A., & Erickson, G. L. (2003). *Teacher inquiry: Living the research in everyday practice.* New York, NY: Routledge.

Clarke, V., & Braun, V. (2017). Thematic analysis. *The Journal of Positive Psychology, 12*(3), 297-298.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9–13.

Conole, G., & Culver, J. (2010). The Design of Cloudworks: Applying social networking practice to foster the exchange of learning and teaching ideas and designs. *Computers & Education, 54*(3), 679–692.

Dawson, K. (2006). Teacher inquiry: A Vehicle to merge prospective teachers' experience and reflection during curriculum-based, technology-enhanced field experiences. *Journal of Research on Technology in Education*, *38*(3), 265–292.

Datnow, A., & Hubbard, L. (2016). Teacher capacity for and beliefs about data-driven decision making: A Literature review of international research. *Journal of Educational Change, 17*(1), 7–28.

Ellis, R. (2008). *The Study of second language acquisition* (2nd ed.). New York, NY: Oxford University press.

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5/6), 304–317.

Gelan, A., Fastré, G., Verjans, M., Martin, N., Janssenswillen, G., Creemers, M., Lieben, J., Depaire, B. & Thomas, M. (2018). Affordances and limitations of learning analytics for computer-assisted language learning: A Case study of the VITAL project. *Computer Assisted Language Learning*, *31*(3), 294–319.

Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American Psychologist, 53*(1), 5–26.

Godwin-Jones, R. (2017). Emerging technologies: Data-informed language learning. *Language Learning & Technology*, *21*(3), 9–27.

Haya, P. A., Daems, O., Malzahn, N., Castellanos, J., & Hoppe, H. U. (2015). Analysing content and patterns of interaction for improving the learning design of networked learning environments. *British Journal of Educational Technology*, *46*(2), 300–316.

Hsiao, I. Y. T., Lan, Y. J., Kao, C. L., & Li, P. (2017). Visualization analytics for second language vocabulary learning in virtual worlds. *Educational Technology & Society, 20*(2), 161–175.

Hu, X. (2017). Automated recognition of thinking orders in secondary school student writings. *Learning: Research and Practice*, *3*(1), 30–41.

Ifenthaler, D., Gibson, D., & Dobozy, E. (2018). Informing learning design through analytics: Applying network graph analysis. *Australasian Journal of Educational Technology*, *34*(2), 117–132.

Johnson, L., Smith, R., Willis, H., Levine, A., & Haywood, K. (2011). *The 2011 horizon report*. Austin, TX: The New Media Consortium.

Laurillard, D. (2013). *Teaching as a design science: Building pedagogical patterns for learning and technology*. New York, NY: Routledge.

Li, H., Flanagan, B., Konomi, S., & Ogata, H. (2018). Measuring behaviors and identifying indicators of self-regulation in computer-assisted language learning courses. *Research and Practice in Technology Enhanced Learning 13*(19), 1–12.

Link, S., & Li, Z. (2015). Understanding online interaction through learning analytics: Defining a theory-based research agenda. In E. Dixon & M. Thomas (Eds.), *Researching language learner interactions online: From social media to MOOCs* (pp. 369–385). San Marcos, TX: CALICO.

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist*, *57*(10), 1439–1459.

Mangaroska, K., & Giannakos, M. N. (2019). Learning analytics for learning design: A Systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies, 12*(4), 516–534.

McKenney, S., & Mor, Y. (2015). Supporting teachers in data-informed educational design. *British Journal of Educational Technology*, *46*(2), 265–279.

Mandinach, E. B., & Jimerson, J. B. (2016). Teachers learning how to use data: A Synthesis of the issues and what is known. *Teaching and Teacher Education, 60*, 452–457. doi:10.1016/j.tate.2016.07.009

Mor, Y., Ferguson, R., & Wasson, B. (2015). Editorial: Learning design, teacher inquiry into student learning and learning analytics: A Call for action. *British Journal of Educational Technology*, *46*(2), 221–229.

Mouri, K., Uosaki, N., & Ogata, H. (2018). Learning analytics for supporting seamless language learning using e-book with ubiquitous learning system. *Educational Technology & Society, 21*(2), 150–163.

Oxford, R. L. (1997). Cooperative learning, collaborative learning, and interaction: Three communicative strands in the language classroom. *The Modern Language Journal, 81*(iv), 443-458.

Papamitsiou, Z., & Economides, A. (2014). Learning analytics and educational data mining in practice: A Systematic literature review of empirical evidence. *Educational Technology & Society, 17*(4), 49–64.

Persico, D., & Pozzi, F. (2015). Informing learning design with learning analytics to improve teacher inquiry. *British Journal of Educational Technology*, *46*(2), 230–248.

Prieto, L. P., Wen, Y., Caballero, D., Sharma, K., & Dillenbourg, P. (2014). Studying teacher cognitive load in multi-tabletop classrooms using mobile eye-tracking. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces* (pp. 339–344). Dresden, Germany: ACM.

Reimann, P. (2016). Connecting learning analytics with learning research: The Role of design-based research. *Learning: Research and Practice, 2*(2), 130–142.

Reinders, H. (2018). Learning analytics for language learning and teaching. *JALT CALL Journal*, *14*(1), 77–86.

Rich, P., & Hannafin, M. (2008). Capturing and assessing evidence of student teacher inquiry: A Case study. *Teaching and Teacher Education*, *24*(6), 1426–1440.

Rodríguez-Triana, M. J., Prieto, L. P., Martínez-Monés, A., Asensio-Pérez, J. I., & Dimitriadis, Y. (2018). The Teacher in the loop: Customizing multimodal learning analytics for blended learning. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 417-426). doi:10.1145/3170358.3170364

Rosé, C. P., Gaesevic, D., Dillenbourg, P., Jo, Y., Tomar, G., Ferschke, O., Yang, S. (2016). Analytics of social processes in learning contexts: A Multi-level perspective. In C. K. Looi, J. Polman, U. Cress, & P. Reimann (Eds.), *Transforming learning, empowering learners: The International Conference of the Learning Sciences*, *Volume 1* (pp. 24–31). Singapore: International Society of the Learning.

Scott, V. M. (1996). *Rethinking foreign language writing*. Boston, MA: Heinle & Heinle.

Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S., & Kirschner, P. A. (2018). Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior, 107*. doi:10.1016/j.chb.2018.05.004

Siemens, G. (2013). Learning analytics: The Emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380–1400.

Song, Y., & Looi, C. K. (2012). Linking teacher beliefs, practices and student inquiry-based learning in a CSCL environment: A Tale of two teachers. *International Journal of Computer-Supported Collaborative Learning, 7*(1), 129-159.

Stapa, S. H., & Majid, A. H. A. (2009). The Use of first language in developing ideas in second language writing. *European Journal of Social Sciences, 7*(4), 41–48.

Reeves, T. D., & Chiang, J. L. (2019). Effects of an asynchronous online data literacy intervention on pre-service and in-service educators' beliefs, self-efficacy, and practices. *Computers and Education, 136*, 13–33.

Voogt, J., Westbroek, H., Handelzalts, A., Walraven, A., McKenney, S., Pieters, J., & De Vries, B. (2011). Teacher learning in collaborative curriculum design. *Teaching and Teacher Education, 27*(8), 1235–1244.

Webster, L., & Mertova, P. (2007). *Using narrative inquiry as a research method.* New York, NY: Routledge.

Wise, A. F., & Vytasek, J. (2017). Learning analytics implementation design. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of Learning Analytics* (pp. 151–160). doi:10.18608/hla17

Wen, Y., Looi, C.K., & Chen, W. (2011). Towards a model for rapid collaborative knowledge improvement in classroom language learning. In H. Spada et al. (Eds.), *International Conference on Computer-Supported Collaborative Learning* (pp. 836–851). Hong Kong, China: International Society of the Learning Sciences.

Wen, Y. (2019). *Computer-supported collaborative Chinese second language learning: Beyond brainstorming*. Singapore: Springer Singapore.

Wong, L. H., Chin, C. K., Chen, W., & Gao, P. (2009). V.S.P.O.W.: An Innovative collaborative writing approach to improve Chinese as L2 pupils' linguistic skills. In *Proceedings of CSCL 2009* (pp. 651–661). Rhodes Island, Greece: International Society of the Learning Sciences.

Yim, S., & Warschauer, M. (2017). Web-based collaborative writing in L2 contexts: Methodological insights from text mining. *Language Learning & Technology*, *21*(211), 146–165.

A Multiple Study Investigation of the Evaluation Framework for Learning Analytics: Instrument Validation and the Impact on Learner Performance

Mehmet Kokoç[1*] and Mehmet Kara[2]
[1]School of Applied Sciences, Trabzon University, Turkey // [2]Vocational School of Design, Amasya University, Turkey // kokoc@trabzon.edu.tr // mehmet.kara@amasya.edu.tr
[*]Corresponding author

(Submitted August 19, 2020; Revised October 15, 2020; Accepted November 16, 2020)

ABSTRACT: The purposes of the two studies reported in this research are to adapt and validate the instrument of the Evaluation Framework for Learning Analytics (EFLA) for learners into the Turkish context, and to examine how metacognitive and behavioral factors predict learner performance. Study 1 was conducted with 83 online learners enrolled in a 16-week course delivered through the Moodle learning management system. The findings from the confirmatory factor analysis indicated that a three-factor model of the EFLA for learners provided the best model fit for the collected data. The model is consistent with the factorial structure of the original instrument developed based on the data from the European learners. Study 2 aimed to reveal how the metacognitive and behavioral factors pertaining to the learning analytics dashboard predict learners' academic performance. A total of 63 online learners enrolled in a 14-week online computing course participated in this study. The results from the logistic regression analysis indicated that online learners more frequently interacted with the learning analytics dashboard demonstrated greater academic performance. Furthermore, the dimensions of the EFLA, together with the interaction with the dashboard, significantly predicted learners' academic performance. This multiple-study investigation contributes to the generalizability of the EFLA for learners and highlights the importance of metacognitive and behavioral factors for the impact of learning analytics dashboards on learner performance.

Keywords: Learning analytics, Learning analytics dashboards, Evaluation, Validation, Learning performance

1. Introduction

Data use in various institutions has been adopted as a way of decision making as a result of the vast amount of data produced through online systems. Learning Analytics (LA) is one of these methods used to make decisions on learning design. LA was defined by Long and Siemens (2011) as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (p. 34). It has emerged as an evolving field of research and practice and has a great potential to transform learning and teaching activities based on data-driven practices (Howell et al., 2018; Ifenthaler & Yau, 2020; Wong & Li, 2020). This potential requires educational researchers and practitioners to pay more attention to LA research and practice. The review studies on LA, however, indicate that the studies generally focus on small-scale implementations (Wong & Li, 2020) and had inadequate empirical evidence on their effectiveness (Larrabee Sønderlund, Hughes, & Smith, 2019; Viberg, Hatakka, Bälter, & Mavroudi, 2018). Thus, the projected potential has not been satisfactorily transformed into practice yet (Viberg et al., 2018) and further research is needed in this regard.

Research shows that the challenges are generally faced during the implementation of the LA interventions such as evaluation of effectiveness (Wong & Li, 2020). For this reason, it is highly desired to continuously evaluate and improve LA interventions and tools (Larrabee Sønderlund et al., 2019; Scheffel, Drachsler, Stoyanov, & Specht, 2014). One of the typically used LA interventions is LA dashboards (Jivet, Scheffel, Drachsler, & Specht, 2017; Jivet, Scheffel, Specht, & Drachsler, 2018). In spite of their common use, little attention was paid to their effectiveness or their impact on learner performance (Verbert et al., 2013; Verbert et al., 2014). It was also underlined that there is a need to further evaluate their effectiveness as pedagogical tools (Jivet et al., 2018). Matcha, Gašević, and Pardo (2020) also point out the restrictions in the way they are evaluated and reported. The current literature, therefore, urges the evaluation of LA dashboards as the pedagogical interventions through the robust evaluation instruments.

In this sense, Viberg et al. (2018) suggest spending more effort on the validation of LA tools and methods. Additionally, the inclusion of the stakeholders such as learners and teachers in the evaluation process is also underlined by the relatively recent studies (Howell et al., 2018; Samuelsen, Chen, & Wasson, 2019; Schumacher & Ifenthaler, 2018). Based on the need for a standardized evaluation framework and instrument, Scheffel et al.

(2014) developed quality indicators for LA and this framework was iteratively implemented, evaluated, and improved (Scheffel, Drachsler, & Specht, 2015; Scheffel, 2017; Scheffel et al., 2017). The final version of the Evaluation Framework for Learning Analytics (EFLA 4), included a solid measurement instrument that enables to gather data from both teachers and learners on the effectiveness of a specific LA tool (Scheffel, 2017). Consequently, the main goal of this multiple-study investigation is firstly to adapt the instrument developed and validated based on this framework and then, to use it to evaluate the impact of an LA dashboard on learner performance. A Prescriptive Learning Dashboard (PLD) was used and its impact on learner performance was evaluated in terms of both metacognitive and behavioral levels. In the present study, metacognitive levels covered the dimensions of the validated EFLA 4 for learners, including "data," "awareness & reflection," and "impact," while behavioral level included dashboard use and the time spent on the dashboard. The specific research questions were as follows:

- How valid and reliable is EFLA 4 for learners in the context of Turkey? (Study 1)
- How do metacognitive dimensions of the EFLA4 instrument for learners and interaction with LA dashboards predict learner performance in an online undergraduate course? (Study 2)

1.1. Evaluation framework for learning analytics

Based on the lack of evaluation standards for LA tools, several versions of the EFLA were created and validated. The EFLA studies aimed to reveal quality indicators for LA tools, to develop and use an instrument based on these indicators, and to validate an instrument to evaluate LA tools (Scheffel, 2017). In this regard, four versions of the EFLA were proposed as EFLA1, EFLA2, EFLA3, and EFLA4, each of which was based on the previous version.

EFLA 1. The first framework was proposed by Scheffel et al. (2014) in the form of quality indicators for LA. The quality indicators were identified as the result of a two-phase study. The participants of the first phase were the stakeholders involved in the implementation of the LA tools while the participants in the second phase were the subject-field experts of the LA. A total of 103 quality indicators were determined and used to create the framework. The first framework covered five areas of quality indicators as the dimensions (Scheffel et al., 2014): (1) "*Objectives*," (2) "*Learning support*," (3) "*Learning measures and output*," (4) "*Data aspects*," and (5) "*Organizational aspects*" (p. 126).

EFLA 2. In their study, Scheffel et al. (2015) evaluated EFLA 1 and used it as a base to construct an instrument to evaluate LA tools. The study was conducted with the participation of the Learning Analytics Community Exchange (LACE) project (http://www.laceproject.eu) and its partners. The study resulted in EFLA 2, with four dimensions for both teachers and learners including three items in each dimension. The identified dimensions for both learners and teachers were labeled as "data," "awareness," "reflection" and "impact" (Scheffel, 2017, p.54). In this way, the first instrument was developed based on EFLA 1 with diminished dimensions and items. According to Scheffel (2017), EFLA 2 met several requirements in the first version, EFLA 1. These were decreasing the number of dimensions and items, facilitating understanding of the dimensions and items, revising the instrument for the purpose of getting answers from both learners and teachers, and grounding the dimensions and items on a theoretical base.

EFLA 3. The previous version was used and evaluated in another second-phase study with the participation of learners and teachers to test its validity and reliability and to identify the problematic aspects of the previous instrument through principal component analysis and focus group interviews with the experts (Scheffel, 2017). The prior instrument was revised and converted into EFLA 3 based on the quantitative and qualitative findings. Thus, EFLA 3 still consisted of four dimensions for both learners and teachers. However, the statements of the items were revised and the number of the items in the dimensions of awareness and reflection was decreased to two. The statements started with "For this LA tool…" or "This LA tool…" instead of "I …".

EFLA 4. Scheffel et al. (2017) used and evaluated EFLA 3 in their two-phase study conducted in a MOOC environment with the participation of both learners and teachers. As a result of the first phase, two items were removed and the instrument has three dimensions for both learners and teachers: "Data," "Awareness & Reflection," and "Impact" (Scheffel, 2017, p.136). The second phase of the validity and reliability analyses indicated that the final version of the instrument, EFLA 4, is an eight-item valid and reliable instrument to evaluate an LA tool from the perspectives of both learners and teachers. Both data and impact dimensions included two items while awareness and reflection dimension included four items (see Table 1). The obtained Cronbach alpha values for the reliability of the instrument were presented in Table 2 for each dimension.

The continuous implementation, evaluation, and revision of the EFLA frameworks resulted in a valid and reliable research instrument to evaluate LA tools (Scheffel, 2017). Thus, the validated instrument could be used in various research contexts to evaluate the effectiveness of LA tools.

Table 1. Dimensions and items of EFLA 4 for learners and teachers

	Item number	Learners	Teachers
Data	1	"For this LA tool it is clear what data is being collected"	"For this LA tool it is clear what data is being collected"
	2	"For this LA tool it is clear why the data is being collected"	"For this LA tool it is clear why the data is being collected"
Awareness & reflection	3	"This LA tool makes me aware of my current learning situation"	"This LA tool makes me aware of my students' current learning situation"
	4	"This LA tool makes me forecast my possible future learning situation given my (un)changed behaviour"	"This LA tool makes me forecast my students' possible future learning situation given their (un)changed behaviour"
	5	"This LA tool stimulates me to reflect on my past learning behaviour"	"This LA tool stimulates me to reflect on my past teaching Behaviour"
	6	"This LA tool stimulates me to adapt my learning behaviour if necessary"	"This LA tool stimulates me to adapt my teaching behaviour if necessary"
Impact	7	"This LA tool stimulates me to study more efficiently"	"This LA tool stimulates me to teach more efficiently"
	8	"This LA tool stimulates me to study more effectively"	"This LA tool stimulates me to teach more effectively"

Note. Retrieved from Scheffel (2017, p. 136).

Table 2. Reliability values for EFLA 4 for learners and teachers

Dimension	*N*	Cronbach alpha (Scheffel et al., 2017)	
		Learners	Teachers
Data	2	.745	.574
Awareness & reflection	4	.916	.870
Impact	2	.954	.881

1.2. Evaluation of learning analytics dashboards

LA dashboards are one of the most commonly used interventions to make decisions about learning design. Schwendimann et al. (2017) define an LA dashboard as follows: "A learning dashboard is a single display that aggregates different indicators about learner(s), learning process(es) and/or learning context(s) into one or multiple visualizations" (p. 37). Based on this definition, data visualization is a key aspect of LA dashboards to inform stakeholders about learning design. LA dashboards work with learner log data, extracting meaning from their data, and visualizing the obtained results (Park & Jo, 2015). Previous studies have indicated their positive effects on learner performance (e.g., Kim, Jo, & Park, 2016; Kokoç & Altun, 2019).

Several review studies on LA dashboards revealed their potential to impact learning design and outcomes (e.g., Bodily & Verbert, 2017; Jivet et al., 2018; Matcha et al., 2020; Schwendimann et al., 2017). One of the earliest review studies was conducted by Verbert et al. (2013). They reviewed 15 learning dashboard applications and proposed a process model for their evaluation. The model includes awareness, self-reflection, sense-making, and impact. By characterizing LA dashboards as the interventions enhancing awareness, reflection, and behavioral change, Verbert et al. (2014) categorized them into three groups: (1) the ones used in face-to-face courses, (2) used in face-to-face groups, and (3) used in online or blended courses. Both of these review studies concluded in common that little research was conducted to evaluate the impact of these dashboard interventions. In their review study, Bodily and Verbert (2017) further argued that there is a need for the dashboard intervention studies focusing not only on the impact, but also on the design and development process through needs analysis, analysis of visual design, and learner surveys. A recent review study by Matcha et al. (2020) similarly concluded that most of the interventions evaluated perceived usefulness while a few of them evaluated the impact.

Although acceptance studies are assumed as a requisite (Jivet et al., 2018), several of the studies underlined that the main research focus is required to be on their impact on learners and learning outcomes (Bodily, Ikahihifo, Mackley, & Graham, 2018; Jivet et al., 2018; Matcha et al., 2020; Park & Jo, 2015; Schwendimann et al., 2017). Jivet et al. (2018) advocated that the primary focus of the research on LA dashboards is required to be on learning goals and the dashboards should be evaluated as the pedagogical tools. Likewise, Matcha et al. (2020) criticized that the studies on LA dashboards are seldom based on learning theory and do not provide suggestions for effective learning. Schwendimann et al. (2017) similarly revealed that more than half of the reviewed studies did not specify any pedagogical approach and suggested evaluation of the interventions in a way that they clearly show the impact on learning. In this sense, Jivet et al. (2018) proposed the levels and criteria for the pedagogical evaluation of the LA dashboards. The levels encompassed metacognitive, cognitive, behavioral, emotional, self-regulation, and usability. Based on these levels and criteria, the current study focused on the evaluation of an LA dashboard at the metacognitive and behavioral levels. Metacognitive level included understanding, agreement, and impact on awareness and reflection while behavioral level included the impact on behavior and system usage (Jivet et al., 2018).

2. Study 1: Scale adaptation and validation

The purpose of Study 1 is twofold: (1) to adapt EFLA 4 for learners into Turkish language and culture and (2) to improve its generalizability through the evidence about its psychometric properties from a different context and participants. In this respect, the present study contributes to the relevant literature on LA by providing a valid and reliable instrument to evaluate the effectiveness of an LA tool. The external validity of EFLA 4 is also to be improved. Thus, this standardized instrument would be used in various contexts to implement, evaluate, and improve LA tools.

2.1. Method

2.1.1. Participants

A total of 83 undergraduate students voluntarily participated in the study. 41 of them (49.4%) were females and 42 of them (50.6%) were males. The participants were the pre-service teachers enrolled in the undergraduate program of "Computer Education and Instructional Technology." All participants stated that they previously registered for at least one fully online course.

2.1.2. Context and procedure

The study was conducted within an undergraduate course entitled "Measurement and Evaluation in Education" at a large-scale public university in Turkey. The content of the course covered the fundamentals of measurement and evaluation in education, validity and reliability, measurement instruments, methods, and item analyses. This 16-week course was delivered through Moodle v3.2 LMS in a semester. In this course, the PLD developed by Kokoç and Altun (2019) was used as an open source LA dashboard compatible with Moodle (see Figure 1). EFLA 4 for learners was distributed as a 10-point Likert scale, ranging from "strongly disagree" to "strongly agree."

The PLD visualizes learners' performance and class average performance in terms of eight LA indicators (basic usage, learning objects, and discussion activities) and assessment scores. Also, the PLD provides personalized real-time recommendations to help learners as text messages based on predictive learner success model.

The scale adaptation procedure was conducted by following the steps recommended by Hambleton (2005). The robustness of the dataset was investigated in the first step through data screening. Missing data, outliers, and floor and ceiling effects were examined in this step. Based on this examination, no data were removed from the dataset as there was no missing data or outlier, and floor and ceiling effects were not observed. The scale items were translated into Turkish and their language equivalency and meanings were tested, and the required items were revised based on the feedback from the subject field experts. Then, the factorial structure of the scale was tested for its validity and reliability in Turkish context.

Figure 1. Screenshots of the used PLD interface (Kokoç & Altun, 2019)

2.1.3. Data analysis

Confirmatory Factor Analysis (CFA) was conducted to examine the construct validity of EFLA 4 for learners. It was also compared with the relevant measurement models. The CFA findings were interpreted based on the fit indices (χ^2/SD, Root Mean Square Error of Approximation - RMSEA, Goodness of Fit Index - GFI, Normed Fit Index - NFI, Standardized Root Mean Square Residual - sRMR, Comparative Fit Index - CFI) as recommended by Jöreskog, Olsson, and Wallentin (2016). Convergent and discriminant validity techniques were also used for construct validity. As for the reliability of the scale and items, item-total correlations, Cronbach Alpha, and Composite Reliability coefficients were computed and evaluated.

2.2. Results

2.2.1. Language validity

The language validity of the scale items was provided through the contribution of the seven professors as experts who have a high level of proficiency in reading, writing and speaking both English and Turkish and have satisfactory knowledge about the literature on the construct of the scale. As the first step, the original items were translated into Turkish by a professor of English language teaching and two professors of instructional technology. The draft form including both Turkish and English items was reviewed by another four-member group (two professors of English language teaching, one professor of Instructional Technology, and one professor of open and distance education) for the appropriateness of the translation through a three-point rubric. Based on the expert evaluations, multi-rater kappa coefficients were computed for each item. The coefficients greater than .60, indicating a good level of consistency, was provided among the experts (Fleiss, 1971). The obtained findings revealed a good level of consistency in terms of language appropriateness as the obtained coefficients for the scale items in the draft form were greater than .80. The scale items were then, implemented with eight undergraduate students for face validity. The findings demonstrated that the scale items were understandable and the scale appeared to measure what it claims to, but the instructions of the scale were required to be detailed more. The draft scale was finalized after the required revisions on the scale.

2.2.2. Measurement model

Figure 2 indicates item-construct parameters of EFLA 4 for learners (standardized factor loadings and the correlations among the factors) obtained through a first-order CFA. Item-construct parameters in Figure 2 show that the standardized factor loadings for the three sub-dimensions of the model range from .65 to .96. The *t* values showed the factor loadings are significant. According to Brown (2015), the factor loadings are required to be greater than .5, and *t* values are required to be significant. The CFA results showed that all goodness-of-fit indices were acceptable (χ^2/ df = .96, p > .05, RMSEA = .01, sRMR = .03, CFI = .98, NFI = .96, GFI = .95), indicating that this eight-item instrument had a good model fit when examined with the data from the Turkish online learners. The results from the CFA revealed that similar findings were gathered with the item-construct structure of the original scale developed with a European sample. This indicated that the construct validity of the adapted scale was of high quality.

2.2.3. Construct validity and reliability

In the study, the construct validity of the scale was determined through convergent and discriminant validity techniques. Convergent validity refers to the degree to which the variables measuring the same construct are associated with each other and the construct they belong (Raykov & Marcoulides, 2011). To provide convergent validity, item loadings obtained for each construct are required to be greater than .05, and the average variance extracted for each construct is required to be equal or greater than .05 and to be less than Cronbach Alpha and composite reliability values (Fornell & Larcker, 1981; Hair, Black, Babin, & Anderson, 2010). According to Nunnally and Bernstein (1994), Cronbach Alpha values are required to be greater than .07. Table 3 shows the average variance extracted, Cronbach Alpha, and composite reliability values.

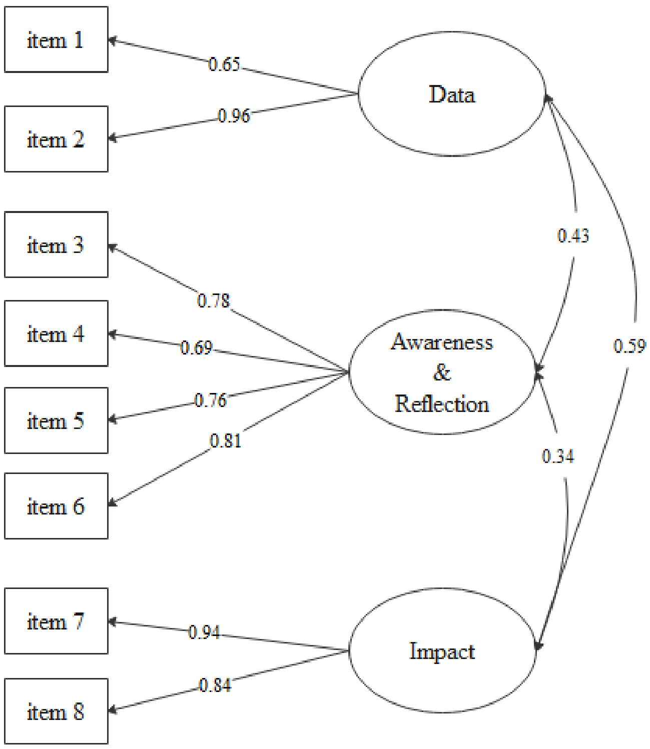


Figure 2. Standardized CFA Solutions for EFLA 4 for learners

Table 3. The average variance values and the reliability coefficients obtained for the constructs

Constructs	Average variance extracted	Composite reliability	Cronbach Alpha (α)
Data	.45	.80	.79
Awareness & reflection	.79	.84	.83
Impact	.58	.88	.89

According to Table 3 the values of the average variance are greater than .05 for the two constructs and less than .05 for one construct. Composite reliability and Cronbach Alpha values obtained for the constructs are greater than .70 as suggested by Nunnally and Bernstein (1994). Although Fornell and Larcker (1981) recommend a value greater than .05 for the average variance explained, they also underline that an obtained composite reliability value greater than .06 for a latent variable is adequate for its convergent validity. Thus, it is concluded that the convergent validity of the scale was obtained as satisfactory and these results indicated the reliability of the total scale is adequate. Discriminant validity refers to the degree to which each latent variable in a measurement model discriminates from others (Farrell, 2010). It requires that the square root of the average variance explained for each construct is not less than the correlation values of each construct with others (Fornell & Larcker, 1981). Table 4 presents the correlations for the constructs in the scale.

Table 4. Correlations among the constructs

Constructs	Data	Awareness & reflection	Impact
Data	$.67^{**}$		
Awareness & reflection	$.43^{*}$	$.89^{**}$	
Impact	$.59^{*}$	$.34^{*}$	$.76^{**}$

Note. $^{**}p < .01$; $^{*}p < .05$.

According to Table 4, correlation coefficients among the constructs are less than the square roots of the average variance values computed for each construct. This finding implies that the discriminant validity of the scale was satisfied. The correlations among the constructs range from .34 to .59 and are moderate and significant.

3. Study 2: EFLA for learners and relation to learner performance

Study 2 aims to investigate the influence of the factors in EFLA 4 for learners and interaction with the dashboard on learners' academic performance. There are various studies in the literature on the influence of the LA dashboards on learner performance. The EFLA instrument for learners developed by Scheffel et al. (2017) measures the metacognitive competencies in relation to LA dashboards. Furthermore, previous studies indicated that behavioral indicators reflecting interaction with LA dashboards can play a key role in evaluating LA dashboards (Kokoç & Altun, 2019; Matcha et al., 2020). In study 2, it was investigated which evaluation constructs together with learner interaction with the dashboard influence learners' academic achievement.

3.1. Method

3.1.1. Participants

The data were collected from 63 out of 66 undergraduate students enrolled in an online course. Their ages ranged from 21 to 26 and the average age was 23.02 (*SD* = 1.75). Three students were excluded from further analysis due to their drop-out from the online course. Of this sample, there were 26 females (41.3%) and 37 males (58.7%). They had taken at least one online/blended course at the university level before the study. The data were collected in the fall term of the 2018 academic year. They are the students of a computer engineering department of a public university and voluntarily participated in the study. All of them were assured as to the confidentiality of their interaction data in the online learning environment.

3.1.2. Context

The study was conducted within a 14-week course entitled "Operating Systems" offered in the third year of a Computer Engineering program at a large university in Turkey. The course was asynchronously delivered four hours per week through Moodle. Several weekly online resources were available such as interactive materials and videos in the learning environment. Weekly assignments and collaborative learning tasks were assigned to the students to reflect upon their knowledge. The PLD was embedded in the Moodle as an LA dashboard. Interaction data reflecting all clicking events of the students were recorded automatically.

3.1.3. Data collection and analysis

The data about the learners' evaluation of the PLD were collected by using the "EFLA 4 for learners" instrument, adapted to Turkish in Study 1. The instrument was distributed to learners as paper-and-pencil, and as a 10-point Likert scale, ranging from "strongly disagree" to "strongly agree." Learners' completion of the scale took about three minutes. The student-generated trace data derived from Moodle time-stamped logs were used to explore learner interaction with the PLD. The interaction data were extracted from the Moodle database by using MySQL queries. The raw data were examined through preprocessing and prepared for the data analysis. Firstly, the raw data were transformed into metrics. Logging traces with timestamps, whenever a student opens and closes the PLD, were stored in a Moodle database table titled "pld_log" which was created by the researchers. Time-related data were logged in a PT format. The time format, PT, does not correspond to a numerical type available in the analysis. Therefore, time traces were automatically converted to a UNIX Time format using a formula created in Excel. To find the time spent viewing the PLD, the time difference between the opening time of the event and the closing or transition of another page was calculated. The time difference reflecting "*time spent on the PLD*" was transformed into seconds. The meaning of the "*total view of the PLD*" is that a student opened and displayed the PLD. In addition, as recommended by Kokoç and Altun (2019), the PLD automatic log-out times was set to 120 seconds as the threshold to prevent bias emerged by fake usage.

Three interaction variables were explored as online behavior indicators reflecting learner interaction with the PLD. In total, the study consisted of six independent variables representing metacognitive and behavioral aspects of evaluating LA dashboards as shown in Table 5.

Table 5. Descriptions of Study 2 variables

Aspects	Variables	Items/Description
EFLA 4 for Learners (Metacognitive Aspects)	Data	Item 1: "For this LA tool it is clear what data is being collected" Item 2: "For this LA tool it is clear why the data is being collected"
	Awareness & Reflection	Item 3: "This LA tool makes me aware of my current learning situation" Item 4: "This LA tool makes me forecast my possible future learning situation given my (un)changed behavior" Item 5: "This LA tool stimulates me to reflect on my past learning behavior" Item 6: "This LA tool stimulates me to adapt my learning behavior if necessary"
	Impact	Item 7: "This LA tool stimulates me to study more efficiently" Item 8: "This LA tool stimulates me to study more effectively"
Interaction with the PLD (Behavioral Aspects)	Total number of the PLD views	Total number of times the PLD was opened
	Time spent on the PLD	Total duration of time spent viewing the PLD (minutes)
	Viewing regularity	(1) Not regularly viewing the PLD at least once every week (2) Regularly viewing the PLD at least once every week

The collected data were analyzed based on the research question. The general overview of the dataset was firstly examined through descriptive statistics. No missing value was identified in the dataset. Pearson's correlation analysis was conducted to check the significance of the relationships among the continuous variables. Binary logistic regression was conducted to investigate a predictive model considering the cause-effect relationship between the variables. Logistic regression analysis is a multivariate statistical technique used to compute the probability of the effects of independent variables (predictors) on dependent variables as well as identifying the risk factors (Field, 2013). Logistic regression was used in the current study as it does not require normality and common covariance assumptions, and the categorical variables might be used in the predictive model both as dependent and independent variables. The dependent variable used in the analysis includes two categories: (0) failing the course (unsuccessful), (1) passing the course (successful). The continuous independent variables were the constructs within EFLA 4 for learners, the total number of the PLD views, and the time spent on the PLD. The categorical independent variable was viewing regularity.

Odds ratios (ORs) with 95% Confidence Intervals (CIs) were computed to reveal the probabilities of the sustained attention for each independent variable. The Omnibus test with the model coefficients was used to test the relationships between the combinations of the dependent and independent variables. Nagelkerke R^2 coefficient was used to determine how the independent variables explain the variance of the dependent variables. Hosmer-Lemeshow test was used to investigate the goodness of the model-data fit. During the data analysis, .05 was adopted as the level of significance with the two-tailed tests. Finally, the data collection and analyses were conducted in compliance with the ethical guidelines.

3.2. Results

The correlations between the EFLA instrument scores and the interactions with the PLD in Moodle were presented in Table 6. The table shows that there was a significant correlation between data and impact dimension ($r = .26$, $p < .05$). Also, a significant correlation was found between impact and awareness & reflection dimension ($r = .28$, $p < .05$). Time spent on the PLD was found as significantly correlated with the total view of the PLD ($r = .40$, $p < .05$).

Table 6. Descriptives and correlation analysis results

	Data	Awareness & reflection	Impact	Total view of the PLD	Time spent on the PLD
Data	1				
Awareness & reflection	.18	1			
Impact	$.26^{*}$	$.28^{*}$	1		
Total number of the PLD views	.03	.13	.10	1	
Time spent on the PLD	.13	.07	.23	$.40^{**}$	1
Mean	7.65	7.40	7.23	58.3	99.8
SD	2.25	1.92	2.21	17.5	32.2
Minimum	1	2	2	40	29
Maximum	10	10	10	101	217

Note. $^{**}p < .01$; $^{*}p < .05$.

As for the unique contribution of the study variables in predicting course success (failing the course = 0, passing the course = 1), logistic regression analysis was employed. The independent variables including the EFLA constructs and interaction behaviors with the PLD were included in the logistic regression analysis as the possible predictor variables in the regression model. Table 7 shows the results of the logistics regression analysis.

Table 7. Logistic regression results predicting learner performance

Independent variables			95% CI for Odds Ratio (OR)		
		b (SE)	OR	Lower	Upper
EFLA for Learners (Metacognitive Aspects)	Data	.27(.19)	1.31	0.90	1.92
	Awareness & Reflection	$.65(.25)^{**}$	1.92	1.16	3.18
	Impact	$.67(.27)^{**}$	1.95	1.14	3.31
Interaction (Behavioral Aspects)	Total view of the PLD	$.07(.04)^{*}$	1.08	1.01	1.15
	Time spent on the PLD	.01(.02)	1.01	0.98	1.04
	Viewing regularity of the PLD	-.75(.90)	0.47	0.08	2.77

Note. Reference category for regularity is "1"; $^{**}p < .01$; $^{*}p < .05$.

The results showed that the logistic regression model is statistically significant (-2 log L = 44.17, chi-square = 42.77, $p < .01$). The results from the Hosmer-Lemeshow test revealed that the model has an acceptable fit and the data-model fit is satisfactory (Chi-square = 2.01, $p > .05$). According to the Nagelkerke R^2 values, all of the independent variables account for 66% of the variance in the dependent variable. The classification table shows that the total ratio of the correct classification of the model is 82.5%. In other words, 82.5% of students were correctly classified. Three variables were statistically significant predictors of group membership: Awareness & reflection ($b = .65$, $p < .01$), impact ($b = .67$, $p < .01$), and total view of the PLD ($b = .07$, $p < .05$). It was concluded that the online students who have high awareness & reflection score (OR = 1.92, 95% CI = 1.16-3.18), high impact score (OR = 1.95, 95% CI = 1.14-3.31), more frequently interact with the PLD (OR = 1.08, 95% CI = 1.01-1.15) will more likely successful in the course. As indicated by the odds ratios, membership in the successful group was 1.92, 1.95, and 1.08 times more likely for every one-unit increase in the scores of awareness & reflection and impact dimensions, and total view of the PLD, respectively.

4. Discussion, conclusion, and limitations

This multiple study investigation aimed to adapt and validate the instrument of EFLA 4 for learners into the Turkish context (Study 1), and to examine how metacognitive and behavioral factors predict learners' academic performance (Study 2). The main aim was to explore whether the dimensions of the EFLA for learners and interaction variables were predictive of the extent to which learners completed the online course successfully.

In Study 1, EFLA 4 instrument was adapted to Turkish context. Study 1 appears to be the first research that validates an instrument in the context of evaluating LA dashboards. The results of Study 1 provided strong support for the psychometric qualities of the three-factor, eight-item EFLA 4, including construct validity and reliability. The reliability statistics showed that the instrument had a good level of internal consistency. It was revealed that the Turkish version of EFLA 4 instrument demonstrates the dimensions of the original scale: (1) Data, (2) awareness and reflection, and (3) impact. The CFA results confirmed the three-component structure of EFLA 4 (Scheffel, 2017). Thus, it was revealed that the Turkish version of EFLA 4 is a valid and reliable instrument to evaluate LA dashboards. As a measurement tool, the adapted version of EFLA 4 instrument has been one of the first attempts to thoroughly evaluate LA dashboards in the Turkish higher education context. In the literature, most LA studies have used EFLA 4 to evaluate and compare LA dashboards based on student perceptions (e.g., Broos et al., 2018; Toisoul, 2017). These studies indicate that this instrument is a valuable tool to measure and compare the impact of LA dashboards on educational practices in online and/or blended learning contexts. Thus, it is hoped that adapting the instrument into different cultures could be useful for LA researchers and learning designers from various countries as well as enhancing its generalizability.

In Study 2, we investigated the impact of the dimensions in EFLA 4 for learners and interaction with the PLD on learners' academic performance. The results of Study 2 have led to two issues worthy of further discussion. The first is that the score of "data" dimension did not predict the learner performance although the other EFLA 4 dimensions, "awareness & reflection" and "impact," are the predictors of learner performance. This result may be explained by the relationship between the theory of self-regulated learning and LA. LA plays a key role in helping online learners develop their self-regulated learning skills (Broadbent, Panadero, Lodge, & Barba, 2020). According to the LA process model with four stages based on self-regulated learning (awareness, reflection, sense-making, and impact), LA dashboards visualize learning traces to support awareness, reflection, and sense-making of learners about their learning process (Verbert et al., 2014). If online learners gain insights on their performance through visualized information and reflect them on their learning process, they can change learning behaviors to achieve intended learning goals. Similarly, previous studies indicated that promoting learners' online self-regulation skills via interventions provided by LA dashboards will lead to improved learning performance (Araka, Maina, Gitonga, & Oboko, 2020; Jansen et al., 2019). Thus, our results are in line with those studies focused on the relation between self-regulated learning and LA. The second issue from Study 2 is to discuss the characteristics of PLD with the obtained findings. The PLD used in this study visualizes learners' online behavioral indicators and provides recommendations based on a predictive model. It additionally visualizes the criterion value about a specific indicator expected by a teacher, individual performance of learners, and the average performance of the class on the same graph. It, therefore, has a relatively goal-oriented structure. The results obtained in this study might be the result of these characteristics of the used PLD.

Consistent with the previous studies (Broos et al., 2020; Kokoç & Altun, 2019), Study 2 found that total view of the LA dashboards was a predictor of learning performance and the learners who interacted with the LA dashboard more frequently were also successful in the course more likely. This evidence suggests that students' actual LA dashboard usage is critical in evaluating the impact of LA dashboards. Surprisingly, time spent on the LA dashboard and viewing regularity of the PLD was not found as the predictors of learner performance. This result implies that successful learners pay more attention to the recommendations and interventions offered by the LA dashboard whenever they need them. Another possible explanation for this result is that learners' awareness of the learning process might be enhanced. A learner, who received feedback about his/her learning performance via the PLD, would change his/her learning strategy to improve his/her achievement based on the provided real-time recommendations. For this reason, these results support the conclusion that LA dashboards offering personalized recommendations and interventions influence learner achievement (Ifenthaler & Yau, 2020). In addition, the results should be interpreted with caution because this study is only concerned with one class and one LA dashboard titled PLD. It is important to bear in mind that the results could be different when other LA dashboards are examined in different targets.

The results of Study 2 may make a significant contribution to the institution-wide adoption of LA dashboards and systems. There are some challenges for the implementation of LA initiatives to overcome in higher education (Leitner, Ebner, & Ebner, 2019). Although studies on the design and implementation of LA systems are increasing in the literature, the use of them by large-scale educational institutes worldwide is still at an early stage (Colvin, Dawson, Wade, & Gašević, 2017; Gašević, Tsai, Dawson, & Pardo, 2019). Therefore, it can be inferred that the EFLA instrument for learners is useful for providing evidence about the effects of LA dashboards to educational decision-makers and instructors to facilitate their institutional adoption process and system at higher education institutions. Our results on predicting learner performance based on the EFLA dimensions and interaction with the PLD support this inference. Thus, combining interaction data and self-report data in evaluating the impact of LA dashboards and systems may provide higher education institutions with more valid results and insightful information.

In conclusion, this multiple-study investigation contributes to the generalizability of the EFLA for learners and highlights the importance of both metacognitive and behavioral factors for the impact of LA dashboards on learner performance. In this regard, the current study aimed to address the call for further research on the impact of LA interventions as pedagogical tools on learning (e.g., Jivet et al., 2018; Matcha et al., 2020) and to contribute to the literature by revealing the influence of both metacognitive and behavioral factors on learning performance. The study suggests that evaluation of LA dashboards in terms of metacognitive and behavioral indicators can help learning designers and instructors examine their impact on learners' performance and improve their online learning experience. It is hoped that the insights gained from this study may be of assistance to researchers and learning designers for evaluating the impact of LA dashboards. Even if it is not addressed in the study, it should be kept in mind that qualitative feedback of students may also be important in the evaluation of LA dashboards (Yoo & Jin, 2020).

There are several limitations to this study, which provide directions for future studies. The first limitation is the small sample size. Study 2 was carried out in one online course at a higher education institution. In future studies, large-scale and longitudinal studies can be conducted in various courses with the participation of more learners to test the prediction model of LA dashboards. Further LA studies are needed to confirm the prediction model in other online learning contexts. The second limitation is that learners' emotional and cognitive individual differences were not included in the prediction model. In future studies, students' emotional status when using LA dashboards can be followed using facial expression recognition. Whether students' different cognitive traits and self-regulation levels have a role in evaluating the impact of LA dashboards can be examined in future studies. The final limitation is that this study was conducted in a context where a single LA dashboard was used. Future studies might focus on LA evaluation for learners and instructors in the online learning contexts where multiple LA dashboards were used. The findings from the evaluation of other LA dashboards in various contexts would contribute to the generalizability of the findings of this study.

References

Araka, E., Maina, E., Gitonga, R., & Oboko, R. (2020). Research trends in measurement and intervention tools for self-regulated learning for e-learning environments—Systematic review (2008–2018). *Research and Practice in Technology Enhanced Learning, 15*(1), 1-21. doi:10.1186/s41039-020-00129-5

Bodily, R., Ikahihifo, T. K., Mackley, B., & Graham, C. R. (2018). The design, development, and implementation of student-facing learning analytics dashboards. *Journal of Computing in Higher Education, 30*(3), 572-598. doi:10.1007/s12528-018-9186-0

Bodily, R., & Verbert, K. (2017). Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies, 10*(4), 405-418. doi:10.1109/TLT.2017.2740172

Broadbent, J., Panadero, E., Lodge, J. M., & de Barba, P. (2020). Technologies to enhance self-regulated learning in online and computer-mediated learning environments. In M. J. Bishop, E. Boling, J. Elen, & V. Svihla (Eds.), *Handbook of Research in Educational Communications and Technology: Learning Design* (pp. 37-52). Cham: Springer International Publishing. doi:10.1007/978-3-030-36119-8_3

Broos, T., Pinxten, M., Delporte, M., Verbert, K., & De Laet, T. (2020). Learning dashboards at scale: Early warning and overall first year experience. *Assessment & Evaluation in Higher Education, 45*(6), 855-874. doi:10.1080/02602938.2019.1689546

Broos, T., Verbert, K., Langie, G., Van Soom, C., & De Laet, T. (2018, March). Multi-institutional positioning test feedback dashboard for aspiring students: Lessons learnt from a case study in flanders. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 51-55). New York, NY: ACM. doi:10.1145/3170358.3170419

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford Press.

Colvin, C., Dawson, S., Wade, A., & Gašević, D. (2017). Addressing the challenges of institutional adoption. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of Learning Analytics* (1st ed., pp. 281-289). doi:10.18608/hla17

Farrell, A. M. (2010). Insufficient discriminant validity: A Comment on Bove, Pervan, Beatty and Shiu (2009). *Journal of Business Research, 63*(3), 324-327. doi:10.1016/j.jbusres.2009.05.003

Field, A. (2013). *Discovering statistics using SPSS* (4th ed.). Thousand Oaks, CA: Sage publications.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*, 378-382. doi:10.1037/h0031619

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50. doi:10.2307/3151312

Gašević, D., Tsai, Y.-S., Dawson, S., & Pardo, A. (2019). How do we start? An Approach to learning analytics adoption in higher education. *The International Journal of Information and Learning Technology, 36*(4), 342-353. doi:10.1108/IJILT-02-2019-0024

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). New Jersey, NJ: Pearson.

Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment* (pp. 3-38). Mahwah, NJ: Lawrence Erlbaum Associates.

Howell, J. A., Roberts, L. D., Seaman, K., & Gibson, D. C. (2018). Are we on our way to becoming a "helicopter university"? Academics' views on learning analytics. *Technology, Knowledge and Learning, 23*(1), 1–20. doi:10.1007/s10758-017-9329-9

Ifenthaler, D., & Yau, J. Y. K. (2020). Utilising learning analytics to support study success in higher education: A Systematic review. *Educational Technology Research and Development*, *68*(4), 1-30. doi:10.1007/s11423-020-09788-z

Jansen, R. S., Van Leeuwen, A., Janssen, J., Jak, S., & Kester, L. (2019). Self-regulated learning partially mediates the effect of self-regulated learning interventions on achievement in higher education: A Meta-analysis. *Educational Research Review, 28*, 1-20. doi:10.1016/j.edurev.2019.100292

Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017). Awareness is not enough: Pitfalls of learning analytics dashboards in the educational practice. In *European Conference on Technology Enhanced Learning* (pp. 82-96). Springer, Cham. doi:10.1007/978-3-319-66610-5_7

Jivet, I., Scheffel, M., Specht, M., & Drachsler, H. (2018). License to evaluate: Preparing learning analytics dashboards for educational practice. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 31-40). New York, NY: ACM. doi:10.1145/3170358.3170421

Jöreskog, K. G., Olsson, U. H., & Wallentin, F. Y. (2016). *Multivariate analysis with LISREL*. Basel, Switzerland: Springer.

Kim, J., Jo, I. H., & Park, Y. (2016). Effects of learning analytics dashboard: Analyzing the relations among dashboard utilization, satisfaction, and learning achievement. *Asia Pacific Education Review, 17*(1), 13-24. doi:10.1007/s12564-015-9403-8

Kokoç, M., & Altun, A. (2019). Effects of learner interaction with learning dashboards on academic performance in an e-learning environment. *Behaviour & Information Technology*, 1-15. doi:10.1080/0144929X.2019.1680731

Larrabee Sønderlund, A., Hughes, E., & Smith, J. (2019). The Efficacy of learning analytics interventions in higher education: A Systematic review. *British Journal of Educational Technology, 50*(5), 2594-2618. doi:10.1111/bjet.12720

Leitner, P., Ebner, M., & Ebner, M. (2019). Learning analytics challenges to overcome in higher education institutions. In D. Ifenthaler, D. K. Mah, & J. Y.-K. Yau (Eds.), *Utilizing learning analytics to support study success* (pp. 91-104). Cham, Switzerland: Springer. doi:10.1007/978-3-319-64792-0_6

Long, P. & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46*(5), 31–40.

Matcha, W., Gašević, D., & Pardo, A. (2020). A systematic review of empirical studies on learning analytics dashboards: A Self-regulated learning perspective. *IEEE Transactions on Learning Technologies, 13*(2), 226-245. doi:10.1109/TLT.2019.2916802

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Park, Y., & Jo, I. H. (2015). Development of the learning analytics dashboard to support students' learning performance. *Journal of Universal Computer Science, 21*(1), 110-133. doi:10.3217/jucs-021-01-0110

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. New York, NY: Routledge.

Samuelsen, J., Chen, W., & Wasson, B. (2019). Integrating multiple data sources for learning analytics—Review of literature. *Research and Practice in Technology Enhanced Learning, 14*(11), 1-20. doi:10.1186/s41039-019-0105-4

Scheffel, M., Drachsler, H., Stoyanov, S., & Specht, M. (2014). Quality indicators for learning analytics. *Educational Technology & Society, 17*(4), 117-132.

Scheffel, M., Drachsler, H., & Specht, M. (2015). Developing an evaluation framework of quality indicators for learning analytics. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 16-20). New York, NY: ACM.

Scheffel, M., Drachsler, H., Toisoul, C., Ternier, S., & Specht, M. (2017). The Proof of the pudding: Examining validity and reliability of the evaluation framework for learning analytics. In *European Conference on Technology Enhanced Learning* (pp. 194-208). Tallinn, Estonia: Springer. doi:10.1007/978-3-319-66610-5_15

Scheffel, M. (2017). *The Evaluation framework for learning analytics*. Heerlen, Netherlands: Open Universiteit. Retrieved from https://core.ac.uk/download/pdf/95759389.pdf

Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior,78*, 397–407. doi:10.1016/j.chb.2017.06.030

Schwendimann, B. A., Rodríguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D., & Dillenbourg, P. (2017). Perceiving learning at a glance: A Systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies, 10*(1), 30-41. doi:10.1109/TLT.2016.2599522

Toisoul, C. (2017). *Design, implementation and evaluation of two widgets for a MOOC's learning analytics dashboard* [Unpublished master's thesis] Open University of the Netherlands.

Wong, B. T. M., & Li, K. C. (2020). A review of learning analytics intervention in higher education (2011–2018). *Journal of Computers in Education, 7*(1), 7-28. doi:10.1007/s40692-019-00143-7

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist, 57*(10), 1500-1509. doi:10.1177/0002764213479363

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., & Klerkx, J. (2014). Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing, 18*(6), 1499-1514. doi:10.1007/s00779-013-0751-2

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The Current landscape of learning analytics in higher education. *Computers in Human Behavior, 89*, 98-110. doi:10.1016/j.chb.2018.07.027

Yoo, M., & Jin, S. (2020). Development and evaluation of learning analytics dashboards to support online discussion activities. *Educational Technology & Society, 23*(2), 1–18.

Matthews, S. D., & Proctor, M. D. (2021). Can Public Health Workforce Competency and Capacity be built through an Agent-based Online, Personalized Intelligent Tutoring System? *Educational Technology & Society*, *24* (1), 29-43.

Can Public Health Workforce Competency and Capacity be built through an Agent-based Online, Personalized Intelligent Tutoring System?

Sarah D. Matthews* **and Michael D. Proctor**
University of Central Florida, College of Engineering and Computer Science, Orlando, FL, USA //
Sarah.Matthews@knights.ucf.edu // Michael.Proctor@ucf.edu
*Corresponding author

(Submitted August 14, 2020; Revised October 26, 2020; Accepted December 5, 2020)

ABSTRACT: The COVID-19 pandemic hit the United States in 2020 resulting in a public health caseload surge precipitating deployment of military and federal medical units, states issuing emergency orders to engage retired medical professionals, and novice or inadequately trained healthcare workers thrust into service to meet the pressing need. The novelty and scope of the pandemic exposed a gap in the competency and the surge capacity of the public health workforce to address the societal needs during the pandemic. This research investigated the capability of an agent-based, online personalized (AOP) intelligent tutoring system (ITS) that adaptively uses aptitude treatment interaction (ATI) to deliver public health workforce training in a prescribed health regime and assure their competency. This research also considers the ability of such an AOP ITS to support rapidly surging capacity of the public workforce to scale to meet healthcare demands while remaining accessible and flexible enough to adapt to changing healthcare guidance. Findings indicate such a system increases participant performance while providing a high level of acceptance, ease of use by users, and competency assurance. However, discussion of our findings indicates limited potential for an AOP ITS using the current ATI paradigm to make a major contribution to adding public health workforce surge capacity unless workforce members are directed to utilize it and technology barriers in the current public health IT infrastructure are overcome.

Keywords: Agent, Intelligent tutoring system, Public health, Surge capacity, Competencies

1. Introduction

In late December 2019, COVID-19 virus struck Wuhan, China (CDC, 2020). In February 2020, the U.S. Centers for Disease Control and Prevention (CDC) indicated COVID-19 was not "spreading in the U.S." (Jernigan, 2020) but by mid-March full on community mitigation phase was initiated (Schuchat, 2020). Epidemiology and surveillance skills, which includes conducting contact investigations and tracing, were needed to protect the society but were deficient since 2000 (Hilliard & Boulton, 2012; Lederberg, 2000). To meet the shortfall, the U.S. deployed Medical Reserve Corp and military and federal medical units, states allowed retired personnel to come back to work, and New York City (NYC) transferred to its public health system nearly 40 experienced contact tracers to lead and supervise 1,000 newly hired contact tracers (State of Florida, 2020; Office of the Mayor, 2020; HHS, 2020). Similar actions happened all over the country (Simmons-Duffin, 2020). Just in time training attempted to fill the gap in knowledge by providing lengthy online documents or providing hands-on experience in a "baptism by fire" strategy (Bauchner & Sharfstein, 2020; HHS, 2020). Since spreading into a full-fledged pandemic across the world, at the time of this writing, COVID-19 has infected more than 17 million people and killed more than 700,000 (JHU, 2020)

Rapid scale up of the public health workforce capacity is critical to prevent, detect, or mitigate an outbreak or pandemic but only if the workforce is competent in their knowledge, skills, abilities (KSA) (ASTHO, 2013; Tao, Evashwick, Grivna, & Harrison, 2018). Underfunding, dependency on categorical funding systems, and the decentralized fragmentation of the United States public health system challenges local and/or state governments to maintain appropriate competency especially in epidemiology (Leider, Coronado, Beck, & Harper, 2018; Soffen & Lu, 2017; Wadman, 2012; Partners, 2018). Budgets allocated for workforce training and development are also volatile. Decreasing by 57% in 2009 (APHA, 2011) and after reversing during restorative Affordable Care Act (ACA) funding have since suffered significant cuts up to 80% (Soffen & Lu, 2017; Wadman, 2012; Yeager, 2018).

1.1. A perspective on training and intelligent tutoring systems (ITS)

Educational programs and information disseminated by the CDC attempt to address public health workforce competency and capacity shortfalls (HHS, 2019). The gold standard for competency is on-the-job pairing of trainees with a seasoned epidemiologist in the field such as the Field Epidemiology Training Program (FETP) and the Epidemic Intelligence Service (EIS) program, which has been successfully implemented in 80 countries and has trained over 18,000 graduates since 1980 (CDC, 2020). On-the-job professional development is difficult (Beck, Boulton, & Coronado, 2014) and loss of talent to competing employers undermines local capacity for expert-to-novice mentorship and tutoring programs (Leider et al., 2018). Educational programs often prioritize licensed medical, dental, and nursing staff missing other licensed and non-licensed staff (Tao et al., 2018). As a result, the Department of Health and Human Services (HHS) estimates that only 20% of public health professionals have the KSA's to be effective (Beck et al., 2014; Hilliard & Boulton, 2012).

Face-to-face, on-the-job training as well as learning from peers within discussion groups competes with online learning approaches (Benta, Bologa, Dzitac, & Dzitac, 2015; Hilliard & Boulton, 2012; Kaur, 2013). On site face-to-face training is not efficient nor cost effective for large, geographically distributed populations with diverse student needs nor even practical in a short time frame due to non-availability of experts to train the less experienced when they are in the midst of addressing a pandemic (HHS, 2020; Sottilare & Proctor, 2012).

Synchronous online learning modalities with live instructors, if available, may be accomplished through distance media such as Zoom, WebEx, etcetera (Kaur, 2013). Asynchronous online learning modalities span independent internet searches of textual materials to online videos, tutorials, and intelligent tutoring systems (ITS) (Kaur, 2013). Proprietary online adaptive ITS systems such as those by educational publisher McGraw Hill and Pearson report antidotal success rates in formal educational institutions (McGraw-Hill, 2020; Pearson Higher Education, 2020). For the public health workforce, it was not until April 2020 that a series of asynchronous online teaching and learning approaches in contact investigations and tracing were introduced by CDC and its partners to respond to the COVID-19 pandemic (CDC, 2020). Learner engagement and competency were not accessed by that online training.

Fischetti & Gisolfi (1990) identified ITS as a computerized system that uses artificial intelligence (AI) to tutor a topic. ITS mimics student teacher interaction by modeling the state of a student learner to provide individual instruction (Ma, Adesope, Nesbit, & Liu, 2014). An ITS requires a knowledgeable domain expert for quality content, but once developed, online ITS can be accessed by many learners (Fischetti & Gisolfi, 1990) achieving greater cost effectiveness than traditional methods through greater scale and reuse (Gurunath, Ravi, & Srivatsa, 2012; Ruiz, Mintzer, & Leipzig, 2006). Online ITS enable any time, any location access that is not limited by classroom capacity (Gurunath et al., 2012). ITS learning management systems allow tracking and monitoring of a learner's KSA's while delivering more standardized course content (Ruiz et al., 2006).

Recognized as effective for many (Ramayah, Ahmad, & Tan, 2012; Yiu & Saner, 2005), newer ITS aim to achieve even better adaption and outcomes by employing agent technology (VanLehn, 2011). Commercial agent technology such as Cortana, Siri, Alexa, and Google Assistant leverage the Internet, vocally respond to questions, and provide information as "smart assistants," though do not currently tutor professional subject matter (Martindale, 2020). To be successful tutors, ITS agents must possess levels of autonomy, responsiveness, reactiveness or adaptability, pro-activeness, and social ability (Wooldridge & Jennings, 1995).

Emerging instructional agent-based, online personalized (AOP) ITS are flexible and adaptive on specialized content and contextually sensitive as a personal human tutor might through aptitude treatment interaction (ATI) (Stoilescu, 2008). ATI adapts learning strategies to specific student characteristics (e.g., prior knowledge, aptitude) in combination with 4 adaptive system modules (domain, learner, pedagogical and tutor-user interface) to provide appropriate feedback and instruction to remediate learning deficiencies through the tutor-user interface presented to the learner (Nguyen & Do, 2008; Sottilare, 2018). Real-time data analysis assesses performance, motivation, engagement, and learning (Fischetti & Gisolfi, 1990; Sottilare, 2018). Advanced ATI's address context and importance of student mood (Sottilare & Proctor, 2012), understanding complex issues and improving decision making (Wolfe et al., 2015), improving motivation (Sottilare, Graesser, Hu, & Goldberg, 2014), and improving training efficiency and flexibility (Oxman & Wong, 2014). Additionally, ITS have grown to span numerous specialized fields including mathematics, physics, and software programming, but not public health (Sottilare, 2018).

1.2. A perspective on usability and usefulness

Clearly a prospective public health trainee will be less inclined to voluntarily use an ITS unless it is perceived as usable and useful.

The Technology Acceptance Model (TAM) (Davis, 1989) mediates constructs of perceived usefulness (PU) - the degree to which a person believes that the use of an application or system will improve their job performance - and perceived ease of use (PEOU) - belief that the use of an application or system would be free of effort. PU and PEOU influence attitude toward use (ATT) and intention to use (IU) which indicates the level of acceptance of the technology (Gefen, Straub, & Boudreau, 2000; Turner, Kitchenham, Brereton, Charters, & Budgen, 2010). TAM is widely used in industries outside healthcare and accounts for 30-40% of IT acceptance assessments (Holden & Karsh, 2008; Legris, Ingham, & Collerette, 2003).

1.3. Competency of public health professional outbreak training

Public health professional competency may be measured in terms of knowledge of and compliance with a prescribed health regime. As an example, a health regime for members of a society combating the COVID-19 pathogen includes knowing to and complying with regular washing of hands with soap for 20 seconds and wearing a facial mask when in a group (CDC, 2020). The U.S. Public Health Service, as early as the 1950's, recognized with the Health Belief Model (HBM) that perception and belief of individuals, even professionals, often challenge compliance with health regimes (Rosenstock, 1974; Champion & Skinner, 2008; Rosenstock, Strecher, & Becker, 1988). HBM hypothesizes that compliance with a prescribed health regime decomposes into four Cues to Action (CA) constructs: perceived susceptibility (PS) to the health threat, perceived severity/seriousness (PSV) of the health threat, perceived benefits (PB) to taking the prescribed action, and perceived barriers/threats (PT). These constructs are combined with the motivation (M) of an individual to undertake the behavior (Utwente, 2017; Rosenstock et al., 1988). As currently seen in the COVID-19 pandemic, varying levels of knowledge and compliance to a prescribed health regime by individuals in the society is reflected in varying influences of these constructs (Clark, Davila, Regis, & Kraus, 2020; Van Bavel et al., 2020). Within the tutor, the prescribed health regime are protocols to manage measles and varicella pathogen. In our study, we are measuring the prescribed health regime as the actual use of an ITS.

In the U.S., the framework most frequently utilized to measure public health professional competency in administering a prescribed health regime is "The Core Competencies for Public Health Professionals." This framework contains 8 domains (PHF, 2014; Tao et al., 2018): Analytical/Assessment Skills, Policy Development/Program Planning Skills, Communication Skills, Cultural Competency Skills, Community Dimension of Practice Skills, Public Health Science Skills, Financial Planning and Management Skills, and Leadership and Systems Thinking Skills (PHF, 2014). Our tutor addresses KSA's in 4 of these domains.

2. Research design and methods

First, this research focuses on training and competency assurance of epidemiology and surveillance skills for public health professionals in administering protocols, procedures and processes for selected outbreak pathogens using an AOP ITS with ATI. Secondly, this research discusses the potential of an AOP ITS with ATI might have toward supporting surging up the epidemiology and surveillance skill capacity of the public health professional workforce needed by a society during an outbreak or pandemic. Thirdly, we utilize two theoretical frameworks (TAM & HBM) to understand perception for actual use of an AOP ITS with ATI by public health professionals.

The study involved an ITS that may address protocols, procedures and processes for COVID-19 or any other pathogen outbreak or pandemic, but for the purpose of this research the ITS addressed measles and varicella pathogens. Specifically, the ITS contained 8 learning concepts for each pathogen (16 total) but the study focused on 4. The concepts include epidemiological information for case and outbreak management. The ITS content is adapted from the Florida Department of Health's Epidemiology and Rash Illness Outbreak Tactics (EPI-RIOT): Combining Epidemiologic Practice with the Field Operations course (FBOE, 2009) and heavily supplemented by information from the CDC. The ITS was built using the Generalized Intelligent Framework for Tutoring (GIFT) platform developed by the Army Research Laboratory (ARL) (Sottilare, 2018).

The research involved survey and testing within a cross-sectional experimental study design engaging invited national, state, and local public health professionals. A pilot validated the ITS, testing, and survey instruments. User feedback from the pilot was used to refine or modify the ITS, the testing scenario, and the survey instruments. The modified course was then evaluated by 3 subject matter experts prior to study deployment.

The tutor course flow can be categorized into 4 sections. Section 1 is administrative and includes a 2-minute course navigation video, informed consent, course expectations, course objectives, 13 question learner attributes survey and a 10-question pre-test assessment with structured review. The structured review is one strategy used to provide feedback for learning. Section 2 contains the adaptive course flow modules for measles and varicella. This section contains the rule content files, example content files and check on learning phases used to assess and remediate learners. In the rule content, all learners are presented with a 9-minute measles overview video and 13-minute varicella overview video prior to the learning phase assessments of each topic. If competence is demonstrated within the learning phase, the learner is moved to the next section. If competence is not achieved, remediation strategies are provided using the example content files based on the learner's input. The example content files are a myriad of PowerPoint presentations, videos, websites, and PDF files. Section 3 contains the knowledge application scenario and a 10-question posttest assessment with a structured review. Section 4 is the research framework surveys which contain 3 questions on platform preferences, 22 questions for TAM and 34 questions for HBM.

The data collection tools are a combination of self-report and objective assessments in free text and multiple-choice formats. These survey tools are used to measure prior knowledge, knowledge acquisition and application, and learner attributes such as grit, motivation, and confidence. Learner remediation can take more time than anticipated by a learner as fundamental concepts may be known but not mastered requiring additional time for course completion. Grit (resilience or perseverance when faced with obstacles) was assessed by confidence in completing the entire course, confidence with the content, and willingness to learn on the ITS. Motivation (the reason for the learner's action) was assessed by asking participants about confidence in completing the entire course, willingness to learn on the ITS, and confidence in returning to the platform for a refresher course (Sottilare et al., 2014). Perception responses were based on TAM developed by Davis (1989) and validated by Davis, Bagozzi and Warshaw (1989) and HBM developed by the Public Health Service (1974) and validated by Champion (1984). Additional data collection methods for each hypothesis is described below.

Informed consent was sought for each phase in compliance with IRB guidance (UCF IRB SBE-18-14393). For both phases, the GIFT platform allows for respondent anonymity.

Our study consists of five hypotheses under test.

- Hypothesis A: Preference for obtaining knowledge
- Hypothesis B: Knowledge acquisition and application
- Hypothesis C: Technology Acceptance Model concepts
- Hypothesis D: Health Belief Model concepts
- Hypothesis E: AOP ITS using ATI attract learners to study on a voluntary basis

Hypothesis A. Our first research question is, "Do public health professionals' prefer an ITS platform, internet search, mentor or discussion group training modality?". The null hypothesis is that public health professionals are ambivalent about training modality. For learner perceptions on the ITS, the 3-question comparative analysis survey used Yes and No responses. The data collection for Hypothesis A is contained within a survey in section 4 of the tutor course flow.

Hypothesis B. Our second research question is: "Does an AOP ITS that uses ATI improve a public health professionals knowledge level and application of knowledge in an outbreak scenario?" The two-part null hypothesis is that the AOP ITS with ATI will not demonstrate participants improved post-assessment performance level over pre-assessment performance level or competency in applying knowledge in an outbreak scenario assessment.

Learner's knowledge improvement was assessed by survey evaluation of pre (given at the beginning of the course) and post (given at the end of the ITS instruction) performance. Prior knowledge focused on knowledge and experience with the health regime for a febrile rash illness and packaging and shipping clinical specimens. The Brenner's Novice to Expert model was used in the learner attribute survey to understand respondent's level of expertise. The model is composed of domains that differentiates theoretical knowledge from practical knowledge for clinical practice competencies. Brenner's clinical competency scale includes: (1) Novice = Minimal or only textbook knowledge; (2) Beginner = Some working knowledge; (3) Competent = Good background knowledge and area of practice; (4) Proficient = Depth of understanding of discipline and area of practice; (5) Expert = Comprehensive and authoritative knowledge (Kak, Burkhalter, & Cooper, 2001).

Knowledge application (competency) was evaluated in the ITS with an assessment requiring the learner to apply knowledge obtained to a scenario. It is also applied at the conclusion of each learning module wherein, a 4-question assessment is presented addressing the 4 learning concepts. Performance on these assessments adapts the tutor to move forward with the course if the learner demonstrates mastery of the concepts. If mastery is not obtained, the ITS re-formulates content delivery or medium and the learner is remediated until the criteria is met. Data collection for Hypothesis B is contained within the survey and assessment tools in sections 1, 2, and 3 of the tutor course flow.

Hypothesis C. Our third research question is, "Does an AOP ITS that uses ATI promote senses of useful, easy to use, positive attitude, and intention to use in public health professional users?" The null hypotheses are that public health professionals will be ambivalent about the usefulness (PU), ease of use (PEOU), attitude (ATT), or intent to use (IU) an AOP ITS with ATI. Learner's perception of the ITS was recorded for technology acceptance using a Likert 7-point scale (Table 5). Data collection for Hypothesis C is contained within the survey tools in section 4 of the tutor course flow.

Hypothesis D. Our fourth research question is, "Does content in an AOP ITS that uses ATI communicate perceived susceptibility, severity, threat, benefit, cue to action or motivation in public health professional users for the selected outbreak pathogen or prescribed health regime?" The null hypothesis is that public health professional users of the AOP ITS with ATI will be ambivalent about the perceived susceptibility (PS), severity (PSV), threat (PT), benefits (PB), cues to action (CA) or motivation (M) toward the selected pathogen or prescribed health regime. Like the TAM, the HBM used a Likert 7-point response scale (Table 7). Data collection for Hypothesis D is contained within the survey tools in section 4 of the tutor course flow.

Hypothesis E. Our fifth research question is, "Does an AOP ITS using ATI attract invited public health professionals to receive public health professional's knowledge and application to meet a pathogen outbreak scenario?" The null hypothesis is that public health professionals will not voluntarily engage in non-mandatory training for the given pathogen outbreak scenario. This hypothesis is addressed through the response level of invited public health professionals to partake in various training stages. Data collection for Hypothesis E was conducted utilizing a non-participation survey tool presented in the recruitment invitation composed in Qualtrics (Qualtrics, 2020).

Data was extracted from the GIFT and Qualtrics platforms. Microsoft Excel and IBM SPSS Statistics software package were used in data analysis (IBM, 2018). The Wilcoxon signed-ranks test is a non-parametric equivalent of the paired *t*-test. It does not assume normality in the data and is used to compare paired observations by testing difference in mean or median. In our analysis we utilize the median. There are 3 assumptions that must be met to use the Wilcoxon Signed Ranks Test. The first assumption is that your dependent variable is measured at the ordinal or continuous level. Our data utilizes 7-point Likert items (Tables 5 and 7). The second assumption is that the dependent variable should consist of two categorical related groups or matched pairs. We utilize the same study participants for the pre- and post-assessment evaluations. The third assumption is that the distribution of the differences between the two related groups needs to be symmetrical in shape (LAERD, 2018; Influential Points, 2020). In our hypothesis testing methods, our One-Sample Wilcoxon-Signed Rank test value was "4" which represents "neither disagree or agree" and "neither" in the response scales for TAM and HBM, respectively. We utilized power at $\alpha = 0.05$ and $\beta = 0.4$ with confidence intervals of 95%. Significance and decision results are shown in Tables 6 and 8.

3. Results

The pilot study contained 2 focus groups totaling 17 public health professionals from two local county health departments, all with varied experience in surge events. Focus group sessions had overwhelming positive responses and did identify and resolve some questionnaire and computer technical issues. More importantly, discussion revealed that ATI remediation resulted in excessively longer length of ITS training than expected by some participants but also revealed gaps in user knowledge on using an ITS cloud-based platform. For the formal study, detailed information about how the ITS adaptively used ATI for remediation were communicated to participants through recruitment documents. A step-by-step user document (outside the ITS) and an ITS course navigation video demonstrated ITS tools and how to navigate within the course. Respondents also had the ability to contact the researchers to troubleshoot technology barriers or they could reply to the non-participation survey with the choice of "information technology barrier (i.e., system compatibility)."

Participation in the study was voluntary and resulted in the following number of participants at each stage: 940 invitations were sent to national, state and local public health professionals, 179 made course queries, 129 signed informed consents, 104 completed learner attributes surveys, 97 completed pre-test assessments, 73 completed the course and application scenario question, 72 completed the post-test assessment, and 69 completed the technology acceptance model survey and the health belief model survey. There were 42 participants who did not make a course query but did complete a non-participation survey discussed below.

The 69 participants that completed the course and the surveys in their entirety form the cohort used to address hypotheses A through D. Our study population profile mirrored the results of other public health workforce studies as displayed in Table 1 (Jones, Banks, Plotkin, Chanthavongsa, & Walker, 2015). 68% of participants reported experience with content contained in the ITS which is practical for remediated instruction.

Table 1. Demographic data study of participant cohort (n = 69)

Mean age (yrs.)	Age range (yrs.)	Gender (n)		Mean experience (yrs)	Experience range	Experience with Rash Illness (n)	
		Female	Male			Female	Male
43.7	24-69	52 (75%)	17 (25%)	15.7	1-45	47 (68%)	22 (32%)

Note. Demographic data of the study population mirrors the results of other public health workforce studies.

Hypothesis A: Preference for obtaining knowledge. In a comparative analysis, we asked participants if time would have been better spent on researching the content on the internet, talking with a knowledgeable mentor or taking a class with a discussion group rather than taking the course on the ITS platform. The ITS platform was significantly preferred over the 3 choices (Table 2).

Table 2. Comparison-Time would have been better spent with Internet search, Knowledgeable Mentor or Class Discussion Group rather than ITS (n = 69)

Response	Internet search (n)	Knowledgeable mentor (n)	Class discussion group (n)
Yes	11(15.9%)	25 (36.2%)	18 (26.1%)
No	58 (84.1%)	44 (63.8%)	51 (73.9%)

Note. Comparative analysis of methods for obtaining content contained in the ITS platform demonstrates that the ITS is consistently preferred over the 3 methods presented.

Hypothesis B: Knowledge acquisition and application. The competency level of study participants using an ITS and packing and shipping clinical specimens for rash illness are at the lower end of the Brenner Scale while the competency in managing a patient with rash illness shows equality across novice, competent and proficient categories (Table 3).

Table 3. Competency Level of using an ITS, managing a patient with rash illness, packing, and shipping clinical specimens for rash illness (n = 69)

Skill	Novice (n)	Beginner (n)	Competent (n)	Proficient (n)	Expert (n)
ITS	45 (65.2%)	15 (21.7%)	6 (8.7%)	3 (4.3%)	0
Pt Rash	19 (27.5%)	9 (13.0%)	19 (27.5%)	19 (27.5%)	3 (4.3%)
Pack/Ship	25 (36.2%)	17 (24.6%)	19 (27.5%)	6 (8.7%)	2 (2.9%)

Note. Self-reported competency levels of study participants on the Brenner scale for using an ITS, packing and shipping clinical specimens for rash illness and for managing a patient with rash illness.

The average test scores for the pretest was 6.8 points or 68%, the average for the post test was 8.7 points or 87% (p < .01). The descriptive statistics show that there is an increase in scores from pre to post tests. The 25th percentiles saw an increase of 2 points, the 50th by 2 points and the 75th percentile by 2 points. The test statistics show that the ITS indeed demonstrates a statistically significant change in learning effectiveness (Z = -6.05, p < .01). There was a 288% increase for respondents to receive all 10 points and a 150% increase for respondents to receive 9 points. 20% of respondents improved their post test scores by 2 points, 19% by 3 points, 17% by 1 point, 10% by 4 points, 4% by 5 points, 1% by 6 points, and 1% by 7 points. Seventeen 17% percent (N = 12) of respondents did not show any increase or decrease in points when comparing their pretest to their post test scores. Seven percent 7% (N = 5) of respondents showed a decrease of 1 point and 1% (N = 1) a decrease of 2 points (Table 4). In the knowledge application scenario, 75% (52/69) of respondents were able to demonstrate their ability to apply the knowledge gained (Table 4).

Table 4. Scoring of assessments

Points	Pre-test (*n*)	Post test (*n*)	Change in points (*n*)	% Change from Pre to Post	Scenario (*n*)
-2	N/A	N/A	1 (1%)		
-1	N/A	N/A	5 (7%)		
0	0	0	12 (17%)	0	17 (25%)
1	0	0	12 (17%)	0	52 (75%)
2	1 (1%)	0	14 (20%)	-100%	
3	2 (3%)	2 (3%)	13 (19%)	0%	
4	9 (13%)	2 (3%)	7 (10%)	-78%	
5	4 (6%)	0	3 (4%)	-100%	
6	10 (14%)	1 (1%)	1 (1%)	-90%	
7	15 (22%)	6 (9%)	1 (1%)	-60%	
8	14 (20%)	12 (17%)	0	-14%	
9	6 (9%)	15 (22%)	0	150%	
10	8 (12%)	31 (45%)	0	288%	

Note. Based on the scoring assessment data, an AOP ITS using ATI supports skill and competency training for public health professionals.

Hypothesis C: Technology Acceptance Model concepts. Learner perception levels for the TAM concepts were measured using the scale shown in Table 5. Results are graphically displayed in Figure 1. Inferential comparisons of TAM concepts to ambivalence of use are displayed in Table 6.

Table 5. Technology Acceptance Model response scale

1	2	3	4	5	6	7
Extremely Disagree	Disagree	Slightly Disagree	Neither disagree or agree	Slightly Agree	Agree	Extremely Agree

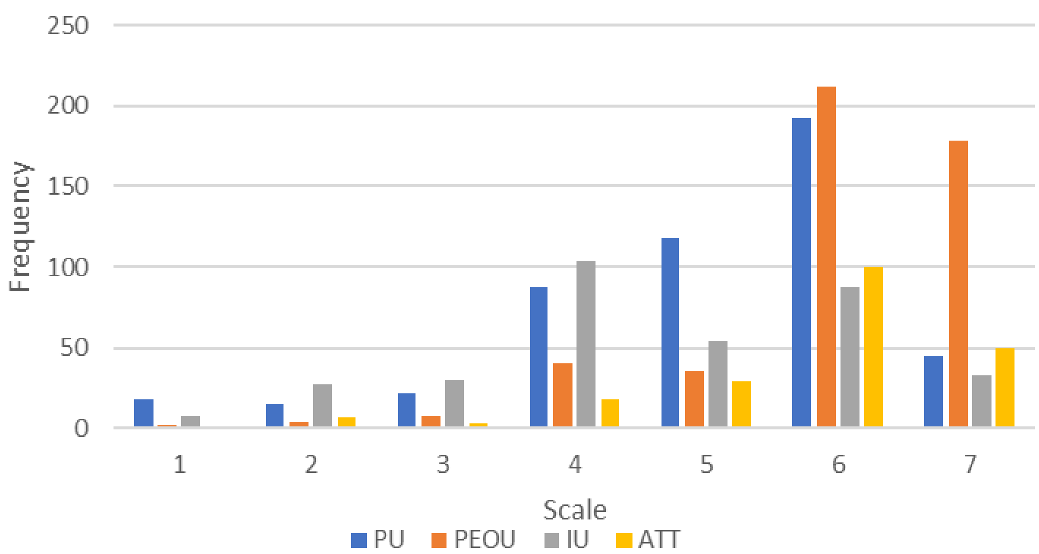


Figure 1. TAM Aggregate Data (n = 69).

Results in Figure 1 and Table 6 indicate that public health professionals are not ambivalent but rather in agreement in using an AOP ITS as it correlates to PU, PEOU and ATT as the mode of their responses on each concept was "Agree." However, there is a level of ambivalence in IU particularly in the temporal indicators for future use (i.e., over the next 3 months).

Table 6. One-sample Wilcoxon Signed Rank Test TAM concepts

Model concepts	Indicators	Label	$\alpha = 0.05$	H_0	$\beta = 0.4$	H_0
Attitude						
	Good idea to Use	ATT1	$p < .01$	Reject	$p < .01$	Reject
	I like the idea to Use	ATT2	$p < .01$	Reject	$p < .01$	Reject
	Using it is a pleasant experience.	ATT3	$p < .01$	Reject	$p < .01$	Reject

Perceived ease of use						
	Easy to Operate	PEOU1	$p < .01$	Reject	$p < .01$	Reject
	Easy to do what I want it to do.	PEOU2	$p < .01$	Reject	$p < .01$	Reject
	Interaction was clear and understandable	PEOU3	$p < .01$	Reject	$p < .01$	Reject
	Flexible to interact with.	PEOU4	$p < .01$	Reject	$p < .01$	Reject
	Easy to become skillful at using	PEOU5	$p < .01$	Reject	$p < .01$	Reject
	Overall, easy to use	PEOU6	$p < .01$	Reject	$p < .01$	Reject
	Over the last 12 months, easy to use.	PEOU7	$p < .01$	Reject	$p < .01$	Reject
Perceived usefulness						
	Enable to accomplish tasks more quickly.	PU1	$p < .01$	Reject	$p < .01$	Reject
	Improve my job performance	PU2	$p < .01$	Reject	$p < .01$	Reject
	Increase productivity.	PU3	$p < .01$	Reject	$p < .01$	Reject
	Enhances effectiveness on the job	PU4	$p < .01$	Reject	$p < .01$	Reject
	Easier to do my job.	PU5	$p < .01$	Reject	$p < .01$	Reject
	Overall, useful in my job.	PU6	$p < .01$	Reject	$p < .01$	Reject
	Over the last 12 months, useful in job.	PU7	$p < .01$	Reject	$p < .01$	Reject
Intention for use						
	Intend to use it for training.	IU1	$p < .01$	Reject	$p < .01$	Reject
	Predict will use it for training.	IU2	$p < .01$	Reject	$p < .01$	Reject
	Expect to use it.	IU3	$p < .01$	Reject	$p < .01$	Reject
	Over the next 3 months, expect to use	IU4	$p = 0.183$	Retain	$p = 0.183$	Retain
	Over the next 3 months, intend to use	IU5	$p = 0.91$	Retain	$p = 0.91$	Retain

Note. Public health professional ambivalence levels toward "Attitude," "Perceived Ease of Use," "Perceived Usefulness" and "Intention for Use." Ambivalence could NOT be rejected for the following dimensions of "Intention for Use": Over the next 3 months, expect to use and Over the next 3 months, intend to use.

Hypothesis D: Health Belief Model concepts. Learner perception levels for the HBM concepts were measured using the scale shown in Table 7. Results are graphically displayed in Figure 2. Inferential comparisons of HBM concepts to ambivalence of use are displayed in Table 8.

Table 7. Health belief model response scale

1	2	3	4	5	6	7
Extremely Unlikely	Unlikely	Slightly Unlikely	Neither	Slightly Likely	Likely	Extremely Likely

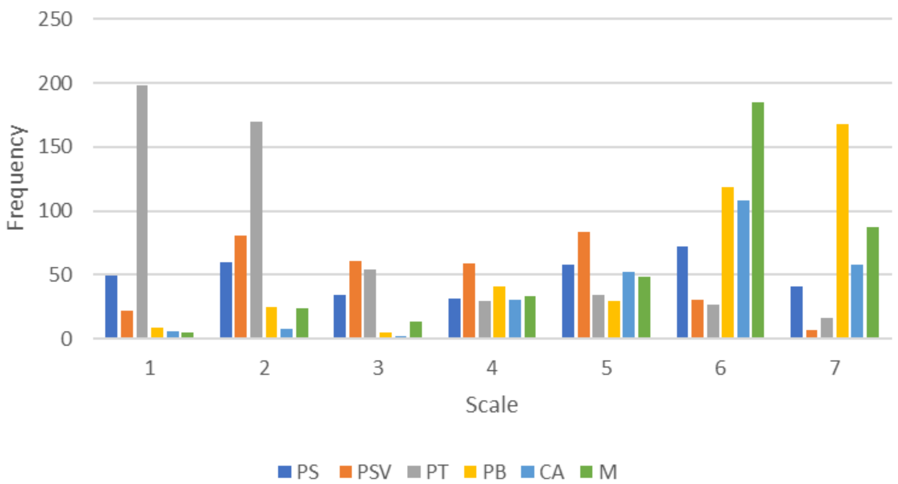


Figure 2. HBM Aggregate Data ($n = 69$)

Table 8. One-Sample Wilcoxon Signed Rank Test HBM Concepts

Model concepts	Indicators	Label	$\alpha = 0.05$	H_0	$\beta = 0.4$	H_0
Perceived susceptibility	Chances of getting a febrile rash illness	PS1	$p < .01$	Reject	$p < .01$	Reject
	Chance of community febrile rash illness outbreak in the future	PS2	$p < .01$	Reject	$p < .01$	Reject
	Likelihood community exposure to an outbreak	PS3	$p < .01$	Reject	$p < .01$	Reject
	Over last 12 months, myself susceptible to a febrile rash-like illness.	PS4	$p < .01$	Reject	$p < .01$	Reject
	Over last 12 months, community susceptible to rash illness outbreak	PS5	$p < .01$	Reject	$p < .01$	Reject
Perceived severity	Over the last 12 months, severity of infection	PSV1	$p = .007$	Reject	$p = .007$	Reject
	Over the last 12 months, experience long term problems from infection	PSV2	$p < .01$	Reject	$p < .01$	Reject
	Severity of the illness on community	PSV3	$p = .007$	Reject	$p = .007$	Reject
	Community experience long term problems from that outbreak	PSV4	$p = .014$	Reject	$p = .014$	Reject
	Over the last 12 months, community severity of outbreak	PSV5	$p = .41$	Retain	$p = .41$	Retain
Perceived threat	Over the last 12 months, afraid for myself to have the lab testing done	PT1	$p < .01$	Reject	$p < .01$	Reject
	Over the last 12 months, be afraid to perform lab testing for community	PT2	$p < .01$	Reject	$p < .01$	Reject
	I do not know the accurate lab tests required for febrile rash illness.	PT3	$p < .01$	Reject	$p < .01$	Reject
	The laboratory tests required for febrile rash illnesses are not reliable.	PT4	$p < .01$	Reject	$p < .01$	Reject
	Preventing rash illness is next to impossible for myself	PT5	$p < .01$	Reject	$p < .01$	Reject
	Preventing rash illness is next to impossible for the community	PT6	$p < .01$	Reject	$p < .01$	Reject
	Over the last 12 months, threat to myself to be infected	PT7	$p < .01$	Reject	$p < .01$	Reject
	Over the last 12 months, threat to my community to be infected	PT8	$p = .002$	Reject	$p = .002$	Reject
Perceived benefits	Important to know how to stay healthy.	PB1	$p < .01$	Reject	$p < .01$	Reject
	Important that my community knows how to stay healthy	PB2	$p < .01$	Reject	$p < .01$	Reject
	Understanding content decreases chances of exposure for community	PB3	$p < .01$	Reject	$p < .01$	Reject
	Understanding content decreases chances of exposure for myself	PB4	$p = .538$	Retain	$p = .538$	Retain
	Over the last 12 months, training myself will be a benefit to me	PB5	$p < .01$	Reject	$p < .01$	Reject
	Over the last 12 months, training myself benefits my community	PB6	$p < .01$	Reject	$p < .01$	Reject
Cue to action	Gaining more knowledge on a topic would improve confidence	CA1	$p < .01$	Reject	$p < .01$	Reject
	Learning about technology from others influences my use of it.	CA2	$p < .01$	Reject	$p < .01$	Reject
	Learning in a self-paced environment influences my use of technology.	CA3	$p < .01$	Reject	$p < .01$	Reject
	Communication from colleagues about technology influences my use.	CA4	$p < .01$	Reject	$p < .01$	Reject
Motivations	General concern about my health.	M1	$p = .004$	Reject	$p = .004$	Reject
	General concern for health of	M2	$p < .01$	Reject	$p < .01$	Reject

community					
Frequently do things to improve health	M3	$p < .01$	Reject	$p < .01$	Reject
Frequently do things to improve health of community	M4	$p < .01$	Reject	$p < .01$	Reject
Search for new information related to health	M5	$p < .01$	Reject	$p < .01$	Reject
Search for new information related to keeping community healthy	M6	$p < .01$	Reject	$p < .01$	Reject

Note. Public health professional ambivalence levels toward "Perceived Susceptibility," "Perceived Severity," "Perceived Threat," "Perceived Benefits," "Cue to Action" and "Motivations." Ambivalence could NOT be rejected for the following dimensions of "Perceived Severity" and "Perceived Benefits" respectively: Over the last 12 months, community severity of outbreaks and Understanding content decreases changes of exposure for myself.

Results from the Wilcoxon signed-rank test indicate that public health professionals are not ambivalent in using an AOP ITS as it correlates to the HBM concepts of PS, PT, CA, and M. The mode of their responses on the concepts of PS, CA and M was "Likely" but for PT was "Extremely Unlikely". Respondents are not ambivalent for 4 of the 5 indicators for PSV with the mode of "Slightly Likely." The fifth indicator is temporal on the severity of an outbreak on the community and does indicate ambivalence. Respondents are not ambivalent for 4 of 5 indicators for PB with the mode of "Extremely Likely." There is ambivalence on 1 indicator as it pertains to perceived benefits about learning about the content of the ITS to decrease exposure to self.

We further stratified our analysis for PS, PSV, PT and PB in terms of perceptions of self-versus the community. We found for PS the mode toward self as "Unlikely" but toward community as "Likely." For PSV, the mode toward self was "Unlikely" and toward the community was "Slightly Likely." For PT, the mode toward self was "Extremely Unlikely" and toward community was "Unlikely" and for PB "Extremely Likely" for self and for the community.

Our results also revealed that public health professionals are highly influenced to use new technology if they learn about it from others if it is in a self-paced environment and if their colleagues communicate about it to them.

4. Discussion

This article contributes to understanding of adaptive learning with an AOP ITS using ATI hosted on a freely available GIFT platform that could be rapidly created and disseminated to educate the public health workforce in order to stem the adverse effects on a society of a pathogen outbreak or pandemic.

Our research instantiated that an AOP ITS using ATI is a pedagogy suitable for public healthcare professionals. While the content and pedagogy addressed measles and varicella pathogen management, the ITS can deliver content for management of COVID-19 or any other pathogen. Further, the ITS actively tests not only knowledge level but application of knowledge in a scenario that could be applied in an outbreak. Administratively this helps public health professionals confirm skill and competency.

Respondents' perception of susceptibility and severity of illness from rash illnesses was greater for the community than self. Their perceptions of threat for self and community was also improbable as a mechanism for prevention is possible (i.e., vaccination). Respondents agreed that they were motivated to learn about how to keep self and their community healthy and that there were benefits from learning this information on an ITS platform including gaining more confidence in work performance. Being online on a cloud platform, the ITS may deliver training at scale, anytime, anywhere in a cost-effective manner, decreasing the demand for expert human mentors.

The 69 person cohort who completed the course were motivated, found the platform useful and would likely return to it in the future as well as advised that it was preferential when compared to an internet search, class discussion, and even a one-on-one interaction with a knowledgeable mentor. These public health professionals largely agreed that the ITS was useful, easy to use, and had a positive attitude toward its use. The ITS most helped respondents who identified below proficient level of competency on the Brenner scale (i.e., Novice, Beginner) as they demonstrated the greatest learning improvement.

In terms of an outbreak or pandemic, our results reveal that our study population has sufficient motivation that febrile rash illness is relevant, that they believe the community is susceptible to a serious health problem and that the use of the AOP ITS would be beneficial in reducing the threat of illness to the community.

Of the limited published literature in scholarly journals, an ITS typically induces pre to post student learning improvements in the range of 0.25 to 1.0 standard deviation (Kulik & Fletcher, 2016). Our study reports an overall 1.00 standard deviation pre to post improvement for our 69-person cohort signifying significant learning effectiveness using ATI with remediation. For understanding the context of theoretical power and alpha error, volunteer-based sampling is non-probabilistic and therefore there is no formula for computing the required sample size and the traditional N = 30 should suffice (Ritter & Sue, 2007). However, Bujang and Baharum (2016) indicate N = 61 yields R_0 = 0.0, R_1 (alternative hypothesis) = 0.4 for correlation tests with a power of 90% and alpha of 0.05. Cohen (1992) indicates N = 64 detects a mean difference medium effect size (.5 standard deviation) with a power of 80% and alpha of 0.05. The 69-person cohort coupled with a full standard deviation improvement exceeds either recommendation. The 69-person cohort is also favorable considering other published ITS research using only 11 to 58 volunteers for analysis (Davidovic, Warren, & Trichina, 2003; Folsom-Kovarik, Schatz, & Nicholson, 2010; Mcquiggan, Mott, & Lester, 2008).

5. Limitations

A cross-sectional study design has inherent disadvantages as it is designed to capture a specific moment in time which may not be representative of behaviors of our study population over time. It also does not help determine cause and effect very well. We did try to control for these disadvantages by asking temporal questions when it came to usage but our respondent group although willing to use the technology in the future were not able to make affirmative choices for use 3 months into the future.

Respondents identified one significant limitation. Although the course content is taken from the nationally recognized authority on notifiable diseases and conditions, application to the nation may be limited. As with all notifiable conditions it is up to the state to adapt their methods for validation and evaluation (CDC, 2019).

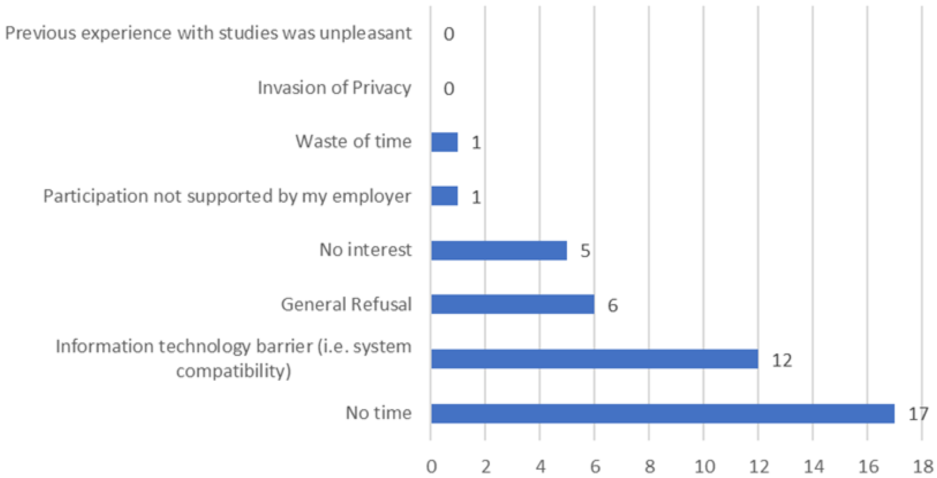


Figure 3. Public health professionals' non-participation in voluntary research (n = 42)

69 participants from 940 invitations indicate limited reach on a voluntary basis of an ITS among the public health workforce. To better understand the 81% non-participation rate, forty-two respondents who did not participate in the study did provide feedback as to why they did not participate (Figure 3). 40% (17/42) identify "no time" and 29% (12/42) identify "information technology barriers (i.e., system compatibility issues)." These two most cited reasons were also validated by email and telephonic discussions. Statistically, Bujang and Baharum (2016) indicate N = 46 yields an R_0 = 0.0, R_1 = 0.4 for correlation tests with a power of 80% and alpha of 0.05. Interpolation of Bujang and Baharum (2016) scale for 42 participants infers a theoretical R1 of .43. Cohen (1992) indicates N = 38 detects a large effect size (.8 standard deviation) mean differences with a power of 80% and alpha of 0.05. 42 respondents coupled with the proportions in two non-participation reasons provide assurance these were the most important reasons for non-participation.

In terms of time, non-participating public health professionals advised that they had too many commitments at work to commit the 30 minutes expected for this research. That infers to reach greater proportions of public healthcare workers, the ITS must be required to be used. Additionally, 179 opened the introduction to the course but the expected 30-minute time for training also proved too optimistic. For the 69-person cohort, not counting one outlier who took 291 minutes to complete the course, the median time of completion of the remaining 68 participants was 46 minutes with a range from 11 to 115 minutes. It is assumed that those spending the greatest amount of time in the system needed the greatest amount of remediation. For the remaining 110, the 30-minute time expectation for the course and the possibility of the course exceeding 30 minutes may explain as much as 2/3rds who did not complete the entire course.

Information technology barriers may also explain as much as 1/3 of the 110 who reneged on completing the course. Specifically, some individuals needed additional instruction on how to connect to the platform and to perform functions within the platform once accessed despite the fore mentioned navigation video explaining connection and use of the platform. More importantly, email communications during the study and the free text responses in the surveys showed that many respondents had course terminations not by their own choice. Many stated that the course “shut down on its own,” would “not allow completion of the process” or would “not move forward or continue.” Later analysis revealed that many health departments do not allow access to cloud applications of this type through their organization firewall. Additionally, many health departments rely on Windows Explorer browsers at their workstations. The prototype used in this research was compatible with Chrome, Edge, or Firefox browsers, not Windows Explorer.

6. Conclusions and future research

Our research indicates that volunteer public health workforce participants completing health regime pathogen outbreak or pandemic training using an AOP ITS with ATI remediation were not ambivalent about the use of the technology. Rather, the synchronous-agent-student engagement facilitated by the ITS design was significantly preferred by participants to an Internet Search, a Mentor, and Classroom discussion. By extension, the inherent scalability, flexibility, and cost effectiveness of the design may also reach but better engage remote learners than the asynchronous e-learning methods currently employed during the COVID-19 pandemic. Further, knowledge increased at a statistically significant amount with volunteers able to effectively apply regimes in an application scenario. Participants perceived the ITS easy to use, useful, and were positively inclined toward it. Additionally, participants were positively inclined toward the ITS particularly toward the HBM concepts of “Perceived Susceptibility,” “Perceived Threat,” “Cue to Action” and “Motivations.” These findings infer the ITS technology may make a significant impact on preparing local and remote workforces to detect, prevent, and respond to public health surge capacity events while providing managers assessment of an individual’s competency with a regime in application scenarios.

Research revealed shortcomings including participants ambivalence about “Perceived Severity,” “Perceived Benefits,” and their “intention to use” the technology in the future. Better understanding the nature of these ambivalences needs future research. These perceptions coupled with the high non-participation rate, forces acceptance of the null hypothesis that for the most part, public health professionals will not voluntarily engage in non-mandatory ITS training for the given pathogen outbreak scenario. By extension, these findings infer limited potential for an AOP ITS using the current ATI paradigm to make a major contribution to adding public health workforce surge capacity unless workforce members are directed to utilize it and technology barriers in the current public health IT infrastructure are overcome.

Findings on time limitations, participant availability limitations, and local constraints also infer future research to better understand how best to address time and availability issues as well as the extent of customization imposed by unique state and organizational governance.

References

American Public Health Association (APHA) (2011). *The Affordable care act’s public health workforce provisions: Opportunities and challenges*. Washington, DC: American Public Health Association. Retrieved from https://www.apha.org/~/media/files/pdf/topics/aca/apha_workforce.ashx

Association of State and Territorial Health Officials (ASTHO) (2013). Public health workforce position statement. *Policy and Position Statements.* Retrieved from http://www.astho.org/Policy-and-Position-Statements/Workforce-Development/

Bauchner, H., & Sharfstein, J. (2020). A Bold response to the Covid-19 pandemic: Medical students, national service, and public health. *Jama, 323*(18), 1790-1791.

Beck, A. J., Boulton, M. L., & Coronado, F. (2014). Enumeration of the governmental public health workforce. *American Journal of Preventive Medicine, 47*(5, Suppl. 3), S306-S313.

Benta, D., Bologa, G., Dzitac, S., & Dzitac, I. (2015). University level learning and teaching via e-learning platforms. *Procedia Computer Science, 55*, 1366-1373.

Bujang, M. A., & Baharum, N. (2016). Sample size guideline for correlation analysis. *World, 3*(1), 37-46.

Centers for Disease Control and Prevention (CDC) (2019). *Nationally notifiable diseases*. Retrieved from https://www.cdc.gov/healthywater/statistics/surveillance/notifiable.html

Centers for Disease Control and Prevention (CDC) (2020). *Coronavirus disease 2019 (COVID-19)*. Retrieved from https://www.cdc.gov/coronavirus/2019-nCoV/index.html

Champion, V., & Skinner, C. (2008). The Health belief model. *Health behavior and health education: Theory, research, and practice*, *4*, 45-65.

Champion, V. L. (1984). Instrument development for health belief model constructs. *Advances in Nursing Science, 6*(3)*,* 73-85.

Clark, C., Davila, A., Regis, M., & Kraus, S. (2020). Predictors of COVID-19 voluntary compliance behaviors: An International investigation. *Global Transitions*, *2*, 76-82.

Cohen, J. (1992). Quantitative methods in psychology: A Power primer. *Psychological Bulletin, 112*, 1155-1159.

Davidovic, A., Warren, J., & Trichina, E. (2003). Learning benefits of structural example-based adaptive tutoring systems. *IEEE Transactions on Education, 46*(2), 241-251.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly,13*(3), 319-340.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A Comparison of two theoretical models. *Management Science, 35*(8), 982-1003.

Department of Health & Human Services USA (HHS) (2019). *CDC's collaboration with academia to strengthen public health workforce capacity*. Retrieved from https://www.grants.gov/web/grants/view-opportunity.html?oppId=301012

Department of Health & Human Services U.S.A. (HHS) (2020). *U.S. Government COVID-19 response plan*. Retrieved from https://int.nyt.com/data/documenthelper/6819-covid-19-response-plan/d367f758bec47cad361f/optimized/full.pdf

Fischetti, E., & Gisolfi, A. (1990). From computer-aided instruction to intelligent tutoring systems. *Educational Technology, 30*(8), 7-17.

Florida Bureau of Epidemiology (FBOE) (2009). *Epidemiology and Rash Illness Outbreak Tactics (EPI-RIOT): Combining epidemiologic practice with field operation*. Tallahassee, FL: Florida Department of Health.

Folsom-Kovarik, J. T., Schatz, S., & Nicholson, D. (2010). Plan ahead: Pricing ITS learner models. In *Proceedings of the 19th Behavior Representation in Modeling & Simulation (BRIMS) Conference* (pp. 47-54).

Gefen, D., Straub, D. W., & Boudreau, M. (2000). Structural equation modeling and regression: Guidelines for research practice. *Communications of the Association for Information Systems, 4*(7), 1-77.

Gurunath, R., Ravi, S., & Srivatsa, S. (2012). Effective professional training through adaptive e-learning and decision-making systems. *International Journal of Scientific & Engineering Research, 3*(11).

Hilliard, T. M., & Boulton, M. L. (2012). Public health workforce research in review: A 25-year retrospective. *American Journal of Preventive Medicine, 42*(5), S17-S28.

Holden, R., & Karsh, B.-T. (2008). The Technology acceptance model: Its past and its future in health care. *Journal of Biomedical Informatics, 43*, 159-172.

IBM Analytics (IBM) (2018). *IBM SPSS Software*. Retrieved from https://www.ibm.com/analytics/data-science/predictive-analytics/spss-statistical-software

Influential Points (2020). *Wilcoxon matched-pairs signed ranks test*. Retrieved from https://influentialpoints.com/Training/wilcoxon_matched_pairs_signed_rank_test-principles-properties-assumptions.htm

Jernigan, D. (2020). Update: Public health response to the Coronavirus disease 2019 outbreak — United States, February 24, 2020. *Morbidity and Mortality Weekly Report, 69*(8), 216-219. Retrieved from https://www.cdc.gov/mmwr/volumes/69/wr/mm6908e1.htm?s_cid=mm6908e1_w

Johns Hopkins University (JHU) (2020). *COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU).* Retrieved from https://coronavirus.jhu.edu/map.html. https://coronavirus.jhu.edu/map.html

Jones, J. A., Banks, L., Plotkin, I., Chanthavongsa, S., & Walker, N. (2015). Profile of the public health workforce: Registered TRAIN learners in the United States. *American Journal of Public Health, 105*(S2), e30-e36.

Kak, N., Burkhalter, B., & Cooper, M.-A. (2001). Measuring the competence of healthcare providers. *Operations Research Issue Paper, 2*(1), 1-28.

Kaur, M. (2013). Blended learning-its challenges and future. *Procedia-Social and Behavioral Sciences, 93*, 612-617.

Kulik, J. A., & Fletcher, J. (2016). Effectiveness of intelligent tutoring systems: A Meta-analytic review. *Review of Educational Research, 86*(1), 42-78.

Laerd Statistics (LAERD) (2018). *Wilcoxon Signed-Rank Test using SPSS Statistics*. Retrieved from https://statistics.laerd.com/spss-tutorials/wilcoxon-signed-rank-test-using-spss-statistics.php

Lederberg, J. (2000). *Summary and assessment-public health systems and emerging infections: Assessing the capabilities of the public and private sectors-workshop summary*. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK100244/

Legris, P., Ingham, J., & Collerette, P. (2003). Why do people use information technology? A Critical review of the technology acceptance model. *Information & Management, 40*, 191-204.

Leider, J. P., Coronado, F., Beck, A. J., & Harper, E. (2018). Reconciling supply and demand for state and local public health staff in an era of retiring baby boomers. *American Journal of Preventive Medicine, 54*(3), 334-340.

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A Meta-analysis. *Journal of Educational Psychology, 106*(4), 901-918.

Martindale, J. (2020). *Cortana vs. Siri vs Google Assistant vs. Alexa*. Retrieved from https://www.digitaltrends.com/computing/cortana-vs-siri-vs-google-now/

McGraw-Hill. (2020). *ALEKS*. Retrieved from https://www.mheducation.com/highered/aleks.html

Mcquiggan, S. W., Mott, B. W., & Lester, J. C. (2008). Modeling self-efficacy in intelligent tutoring systems: An Inductive approach. *User Modeling and User-adapted Interaction, 18*(1-2), 81-123.

Nguyen, L., & Do, P. (2008). Learner model in adaptive learning. *World Academy of Science, Engineering and Technology, 45*(70), 395-400.

Office of the Mayor (2020). *Test and trace: Mayor de Blasio announces New York City test and trace corps*. Retrieved from https://www1.nyc.gov/office-of-the-mayor/news/333-20/test-trace-mayor-de-blasio-new-york-city-test-trace-corps

Oxman, S., & Wong, W. (2014). White paper: Adaptive learning systems. *Integrated Education Solutions*. Retrieved from https://kenanaonline.com/files/0100/100321/DVx_Adaptive_Learning_White_Paper.pdf

Partners in Information Access for the Public Health Workforce (Partners) (2018). Workforce Development. *Education and Training.* Retrieved from https://phpartners.org/educ.html

Pearson Higher Education (2020). *Adapative learning*. Retrieved from https://www.pearsonmylabandmastering.com/northamerica/educators/features/adaptive-learning/index.html

Public Health Foundation (PHF) (2014). *About the core competencies for public health practice*. Retrieved from http://www.phf.org/resourcestools/Documents/Core_Competencies_for_Public_Health_Professionals_2014June.pdf

Qualtrics. (2020). *Qualtrics online survey software*. Retrieved from https://www.qualtrics.com/core-xm/survey-software/

Ramayah, T., Ahmad, N. H., & Tan, S. H. (2012). An Assessment of e-training effectiveness in multinational companies in Malaysia. *Educational Technology & Society, 15*(2), 125-137.

Ritter, L. A., & Sue, V. M. (2007). *The Use of online surveys in evaluation: New directions for evaluation, Number 115*(Vol. 84). San Francisco, CA: Jossey-Bass.

Rosenstock, I. (1974). Historical origins of the health belief model. *Health Education Monographs, 2*(4), 328-335.

Rosenstock, I., Strecher, V., & Becker, M. (1988). Social learning theory and the health belief model. *Health Education Quarterly, 15*(2), 175-183.

Ruiz, J. G., Mintzer, M. J., & Leipzig, R. M. (2006). The Impact of e-learning in medical education. *Academic Medicine, 81*(3), 207-212.

Schuchat, A. (2020). Public health response to the initiation and spread of pandemic COVID-19 in the United States, February 24–April 21, 2020. *Morbidity and Mortality Weekly Report, 69*(18), 551-556. Retrieved from https://www.cdc.gov/mmwr/volumes/69/wr/mm6918e2.htm

Simmons-Duffin, S. (2020). *States get creative to find and deploy more health workers in COVID-19 fight. The Coronavirus crisis*. Retrieved from https://www.npr.org/sections/health-shots/2020/03/25/820706226/states-get-creative-to-find-and-deploy-more-health-workers-in-covid-19-fight

Soffen, K., & Lu, D. (2017). What Trump cut in his budget. *The Washington Post, 16*. Retrieved from https://www.worldpittsburgh.org/wp-content/uploads/2017/06/PDF-What-Trump-Cut-in-His-Budget_The-Washington-Post-March-16-2017.pdf

Sottilare, R., Graesser, A., Hu, X., & Goldberg, B. (2014). *Design recommendations for intelligent tutoring systems Volume 2-Instructional Management* (Vol. 2). Orlando, FL: U.S. Army Research Laboratory.

Sottilare, R., & Proctor, M. (2012). Passively classifying student mood and performance within intelligent tutors. *Journal of Educational Technology & Society, 15*(2), 101-114.

Sottilare, R. (2018). A Comprehensive review of design goals and emerging solutions for adaptive instructional systems. *Technology, Instruction, Cognition & Learning, 11*(1).

State of Florida (2020). *Office of the governor Executive Order Number 20-88 Emergency Management-COVID-19-re-employment of essential personnel, E.O. 20-88 C.F.R.* Retrieved from https://www.flgov.com/2020-executive-orders/

Stoilescu, D. (2008). Modalities of using learning objects for intelligent agents in learning. *Interdisciplinary Journal of E-Learning and Learning Objects, 4*(1), 49-64.

Tao, D., Evashwick, C. J., Grivna, M., & Harrison, R. (2018). Educating the public health workforce: A Scoping review. *Frontiers in Public Health, 6*, 27. doi:10.3389/fpubh.2018.00027

Turner, M., Kitchenham, B., Brereton, P., Charters, S., & Budgen, D. (2010). Does the technology acceptance model predict actual use? A Systematic literature review. *Information and Software Technology, 52*, 463-479.

University of Twente (Utwente) (2017). *Communication studies theories health belief model*. Retrieved from https://www.utwente.nl/en/bms/communication-theories/sorted-by-cluster/Health%20Communication/Health_Belief_Model/

Van Bavel, J. J., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., . . . Druckman, J. N. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, *4*, 460-471.

VanLehn, K. (2011). The Relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197-221.

Wadman, M. (2012). US disease agency in fiscal peril: Proposed budget changes threaten disease prevention and surveillance programmes. *Nature, 483*(7387), 19-20.

Wolfe, C. R., Reyna, V. F., Widmer, C. L., Cedillos, E. M., Fisher, C. R., Brust-Renck, P. G., & Weil, A. M. (2015). Efficacy of a web-based intelligent tutoring system for communicating genetic risk of breast cancer: A Fuzzy-trace theory approach. *Medical Decision Making, 35*(1), 46-59.

Wooldridge, M. J., & Jennings, N. R. (1995). Intelligent agents: Theory and practice. *The Knowledge Engineering Review, 10*(2), 115-152.

Yeager, A. (2018). Cuts to prevention and public health fund puts CDC programs at risk. *The Scientist*. Retrieved from https://www.the-scientist.com/daily-news/cuts-to-prevention-and-public-health-fund-puts-cdc-programs-at-risk-30298

Yiu, L., & Saner, R. (2005). Does it pay to train? ISO 10015 assures the quality and return on investment of training. *ISO Management Systems, 5*(2), 9-13.

Using Gamification to Design Courses: Lessons Learned in a Three-year Design-based Study

Biyun Huang[1*] and Khe Foon Hew[2]
[1]Centre for Learning Sciences and Technologies, The Chinese University of Hong Kong, Hong Kong SAR, China // [2]Faculty of Education, The University of Hong Kong, Hong Kong SAR, China // lucyhuang99@cuhk.edu.hk // kfhew@hku.hk
[*]Corresponding author

(Submitted August 21, 2020; Revised October 15, 2020; Accepted December 5, 2020)

ABSTRACT: A design-based study was conducted in iterative cycles to test the effectiveness of the updated goal-access-feedback-challenge-collaboration (updated-GAFCC) gamification design model. The test-bed was a 10-week undergraduate introductory information management course. Students from three consecutive school years participated in the study, with the control group studying the conventional course without gamification (first year), treatment group_1 studying a gamified course following the original GAFCC model (second year), and treatment group_2 studying an optimized gamified course following the updated-GAFCC model (third year). The results of the design-based study indicated that (i) the updated-GAFCC model and the GAFCC model were effective in enhancing students' learning achievements and task completion; (ii) the updated-GAFCC model was more successful in generating higher quality thinking artifacts than the GAFCC model; (iii) there were fewer lower-quality submissions in the updated-GAFCC condition than in the GAFCC condition; and (iv) 89% of the interviewed students in the updated-GAFCC condition were satisfied with the overall learning design, and felt that the gamified learning activities facilitated their learning. Overall, the findings contribute to our understanding of how pedagogical strategies can be incorporated into the theory-based design model to optimize learning experiences and academic outcomes.

Keywords: Gamification, Design model, Design-based research, Learning performance, Long-term

1. Introduction

Gamification is usually defined as the use of game-like elements, such as badges, leaderboards, and points, in a non-game context (Deterding et al., 2011). Many researchers and practitioners have described gamification as a promising means of motivating students and optimizing their learning outcomes (Bai et al., 2020; Zainuddin et al., 2020). Empirical gamification literature covers K-12 education (e.g., Jong, 2019), university-level education (e.g., Baydas & Cicek, 2019), and short-term online training (e.g., Li et al., 2012). This line of research has investigated the role that gamification plays in behaviorally engaging participants (e.g., Ding, 2019; Barata et al., 2017), improving their academic performance (e.g., Jo et al., 2018), promoting their affective engagement (e.g., Sailer & Sailer, 2020), and catering for learners with different characteristics (e.g., Lopez & Tucker, 2019). In addition, some studies have examined the correlation between the setting of goals and performance in tasks (e.g., Landers et al., 2017).

Overall, the findings of the empirical literature were not consistent, with gamification having been found to be effective in certain situations and ineffective in others. Yildirim (2017) incorporated gamification into a teaching methods course, and reported that students in the gamified course showed higher improvements in learning than the control group. Lo and Hew (2018) found the submission rate of optional assignments in a gamified-flipped mathematics course was higher than that of a traditional course. However, Kyewski and Kraemer (2018) found that the students in the badge conditions did not participate more than the students in the no-badge condition. The inconsistent results suggest that gamification itself does not necessarily improve learning, and that the learning outcomes depend on how gamification is designed and implemented (Dichev & Dicheva, 2017). Thus, we need to explore methods that can optimize the use of gamification in education.

Although design-based research follows an iterative process in systematically examining and refining innovations, and thus produces design principles that can guide similar research efforts (Amiel & Reeves, 2008), this approach is often overlooked in the field of gamification. To date, few studies have used design-based research methods to explore the impact of gamification strategies, especially over a long period (Zainuddin et al., 2020). Moreover, the few empirical design-based studies (e.g., Lee et al., 2013; Cai et al., 2016) reported only the results of the first iterative cycles, not the second cycles, as is conventionally expected. To the best of our knowledge, no design-based studies have iteratively evaluated and refined a theory-based design model that can be used to guide other research studies. To address this gap, our study aims to test and refine a theory-based goal-access-feedback-challenge-collaboration (GAFCC) model using a three-year design-based research approach.

2. Design-based research in gamification

Design-based research seeks to solve real classroom problems that involve collaboration between researchers and practitioners. Design-based studies are typically conducted in iterative cycles of (1) analysis and exploration, (2) design and construction, and (3) evaluation and reflection (Figure 1; McKenney & Reeves, 2012). As shown in Figure 2, these processes facilitate the development of interventions that mature with each iteration. This approach allows the gamification interventions to develop over a longer time and can thus enhance our understanding of the interventions.

Lee et al. (2013) were among the first researchers to use a design-based approach to evaluate the effectiveness of a gamified product. Their study focused on the use of challenges, missions, points, badges, avatars, and leaderboards in the online social system Greenify, which was used to educate adults about issues relating to climate change. The results of their pilot study showed that the participants considered the system to be motivational and fun (79.3%). The system was also found to encourage users to create meaningful, and relevant content for their peers, and increase their awareness of how their behavior and lifestyle might affect the environment (46.2%). The pilot results were quite positive, but no further iterative cycles were conducted.

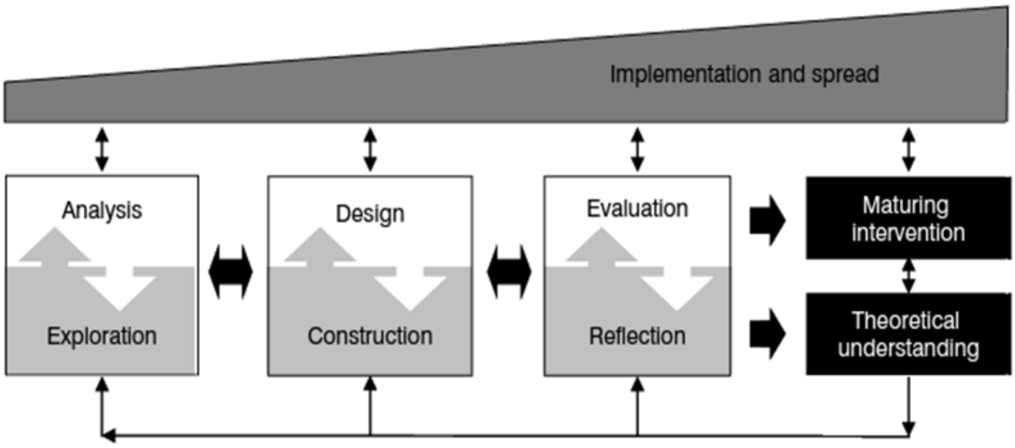


Figure 1. Generic model for design-based research in education (McKenney and Reeves, 2012)

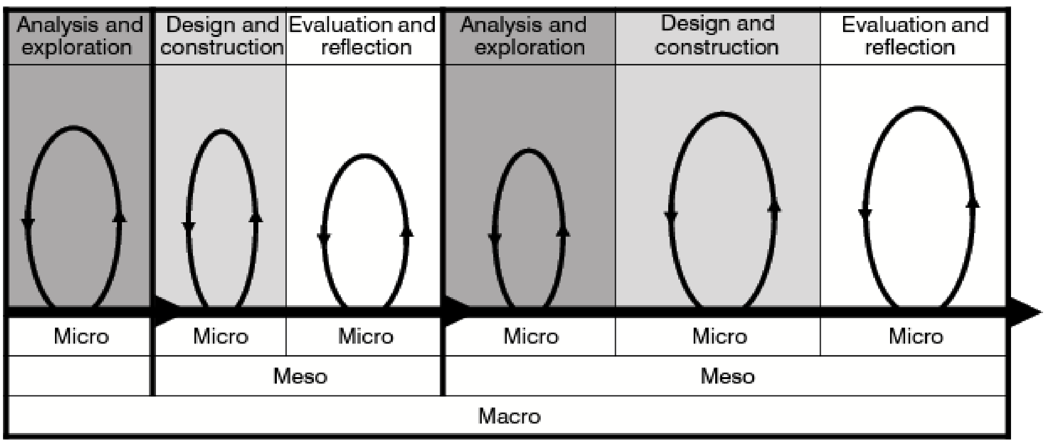


Figure 2. Micro-, meso, and macro-cycles in design-based research (McKenney and Reeves, 2012)

Barata et al. (2017) gamified a multimedia production course and varied their gamification strategies in three iterative cycles. In the first year, they used experience points, leaderboards, badges, and levels. The students were marked on their gamified activities, with the total score contributing to 10% of their final grades. In the second year, the researchers increased the weight of the gamified activities to 20% and added several gamification features, such as the optional Skill Tree and Talkative activities. In the third year, the researchers removed certain quizzes and other challenges to reduce the workload. The results for the third year revealed that the students were more engaged with the activities, and that most students found the course more interesting than other courses. However, the third year students had lower grades (75.4%, *SD* = 14.4) than the first year (79.5%, *SD* = 9.3) and second year students (88.3%, *SD* = 8.1). The results indicated that autonomy and choice, suitable levels of difficulty, and team-based challenges positively enhanced students' participation and satisfaction but did not enhance their learning performance.

Aguilar et al. (2018) iteratively tested a gamification design in a political theory course. In the first round, they allowed the students to choose their assignments, which comprised 60% of the final grade, and power-ups (i.e., awards that students could earn to either increase their grades, make up for absences, or unlock assigned tasks). The first cycle results indicated that the students' perceived autonomy was influenced by the perceived fairness of the grading system and the difficulty of the tasks. The results of the second-round investigation confirmed the initial results. The two rounds showed that gamification supported autonomy and promoted deeper engagement. Overall, the study provided evidence that autonomy is an important element of gamification design.

In summary, although design-based research has helped identify factors (e.g., autonomy) that can improve the motivation of students, none of these studies have contributed to the development and testing of a theory-driven gamification design model. Our study fills this research gap by validating a theory-based meta-model of gamification that will pave the way for future gamification designs.

3. Theoretical framing

This study uses a design-based research framework to evaluate the effectiveness of a meta-model of gamification. The theoretical framework guiding this study is the GAFCC model originally proposed by Huang and Hew (2018). The meta-model consists of two integral parts. The first part introduces the theoretical underpinnings of gamification design and explains how game mechanics can be used to construct a motivational learning environment. The GAFCC model is derived from five motivation theories: goal setting theory (Locke & Latham, 2002), flow theory (Csikszentmihalyi, 1990), self-determination theory (Deci & Ryan, 2000), social comparison theory (Festinger, 1954), and behavioral reinforcement theory (Skinner, 1953). The model highlights that *goal* setting, access to choices and the development of competence, *feedback* on performance and status, (suitable levels of) *challenge*, and a sense of connectedness developed during *collaboration* are important motivational needs. Moreover, specific game mechanics (e.g., badges, leaderboards) corresponding to individual motivational needs can be integrated into learning environments to engage students and encourage expected behaviors (Figure 3).

The second part of the GAFCC model uses the five-step design procedure to gamify a course. This part emphasizes the steps for building alignment between motivational needs, instructional objectives, learning activities, and gamification strategies (Figure 4). We decided to use the GAFCC model as the framework for our design-based study because the model is well grounded in the five motivation theories, and interprets people's motivational needs from the individual and community perspectives. The five-step design procedure is also easy to follow and useful in guiding the gamifying process.

The test-bed was a 10-week undergraduate introductory information management course. The first cycle (pilot study 1) spanned over two years. Students in the first year participated in a non-gamified course. Students in the second year participated in a gamified course based on the GAFCC design model (i.e., treatment group_1). In the second cycle (main study 2), students in the third year participated in a gamified course following the updated-GAFCC design model (i.e., treatment group_2; Figure 5). Pilot study 1 has been reported elsewhere (Huang et al., 2019). In this study, we concentrate on main study 2 and compare the implementation results for treatment group_2 with the results for the control group and treatment group_1. Before describing study 2 in detail, we first summarize the main findings and concerns in pilot study 1.

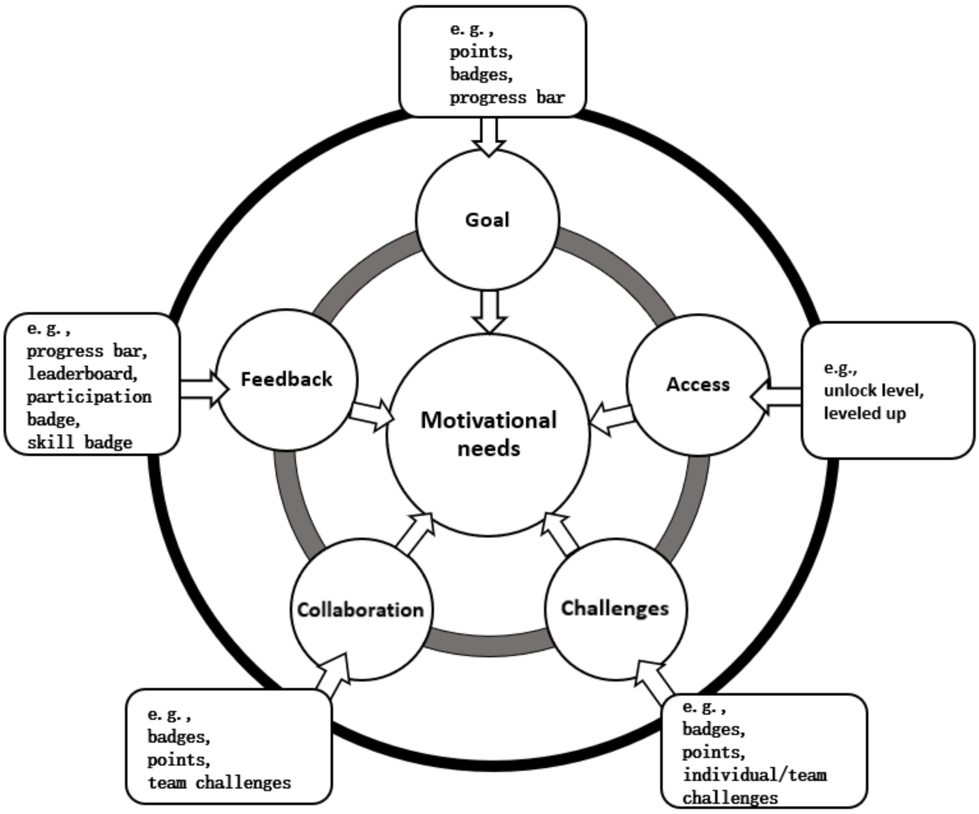


Figure 3. GAFCC model (Huang and Hew, 2018)

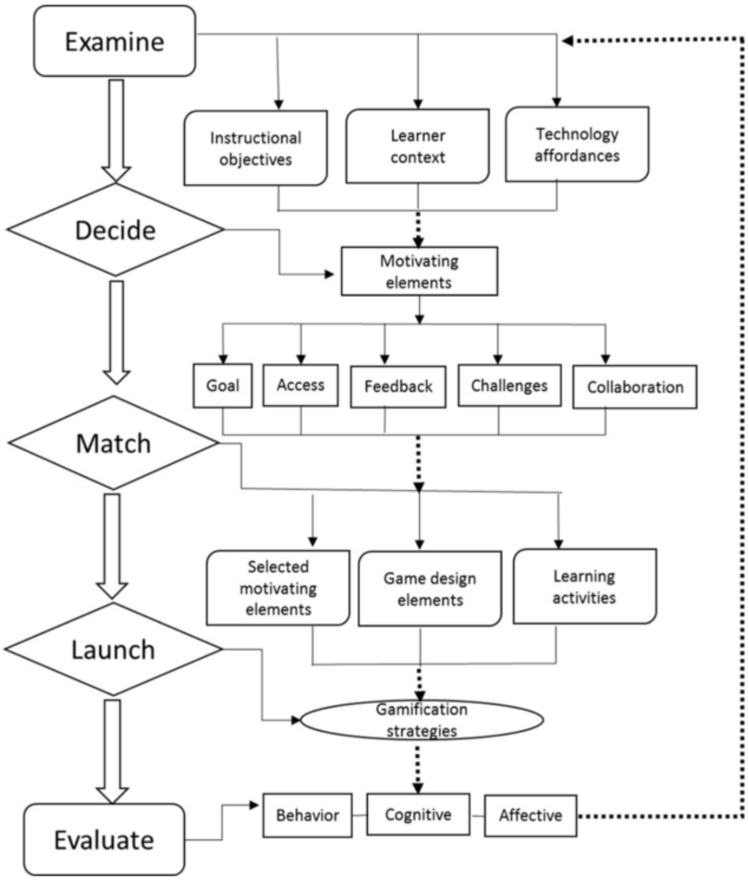


Figure 4. Five-step gamification design procedure (Huang and Hew, 2018)

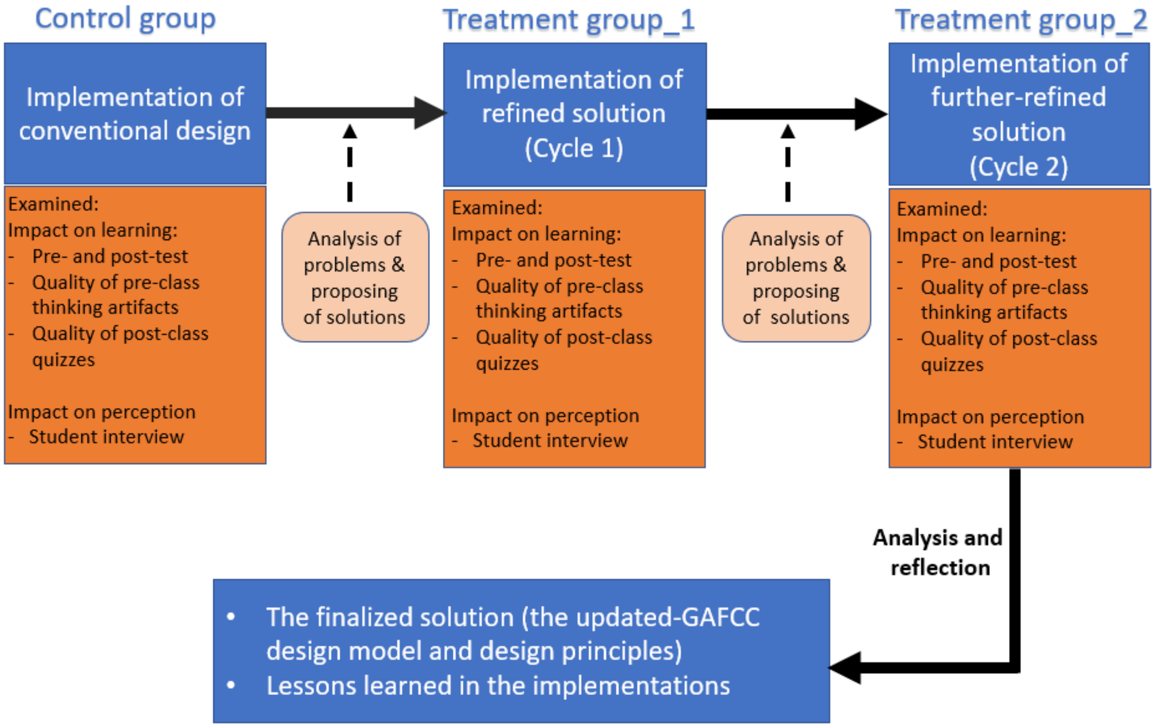


Figure 5. Design of the design-based research project

4. Pilot Study 1: Examining the effects of gamification on student learning and participation

Study 1 was an exploratory pilot study that sought to determine the effects of gamification (if any) on student learning and participation. Forty-eight undergraduates attended an information management course without gamification (control group), while another 48 attended the same course with gamification (treatment group_1). The GAFCC model was used to design treatment group_1 based on the use of badges, levels, and a leaderboard, as shown in Figure 6. Figure 7 shows the different levels used in treatment group_1.

Two types of badges (participation badges and quality badges) were used in treatment group _1. Participation-based badges (e.g., early bird, super-efficient) were used to motivate the students to complete more activities. Quality-based badges (e.g., movie, coffee, tour package, champion) were used to encourage the students to relate their contributions to the contents learned in the course, and provide additional insights on the course-related topics. None of the quality badges could be exchanged for tangible products. The number of badges each student collected was not linked to the course credits.

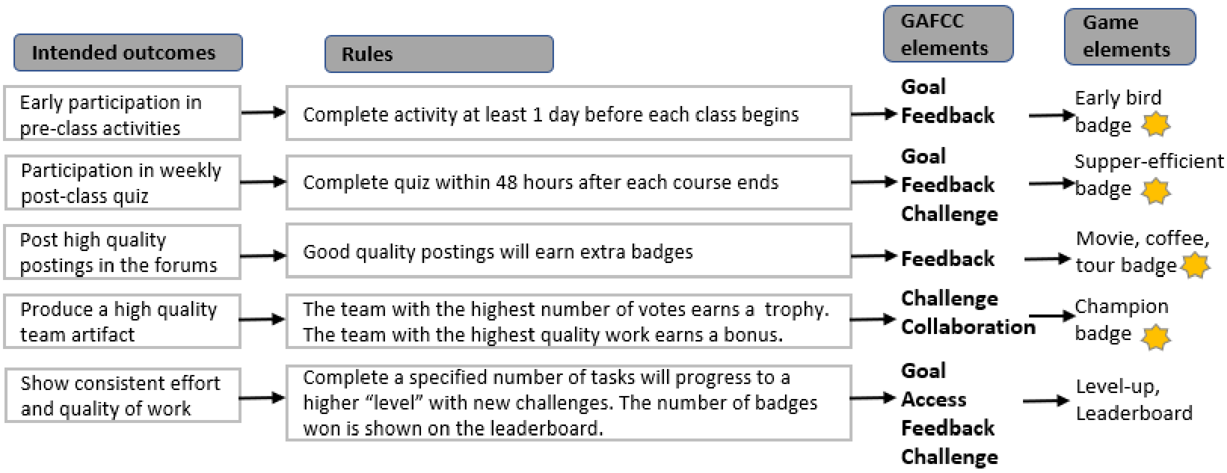


Figure 6. Intended outcomes, rules, GAFCC elements, and game elements in treatment group_1

The results suggested that the GAFCC model was effective in improving student participation and learning. More students in the GAFCC group completed the pre- and post-class activities than students in the control group. In addition, treatment group_1 submitted significantly more in-depth level pre-class thinking artifacts than the control group. Thus, the effectiveness of the positive outcome is, in all likelihood, associated with the fulfillment of the core psychological needs of the students.

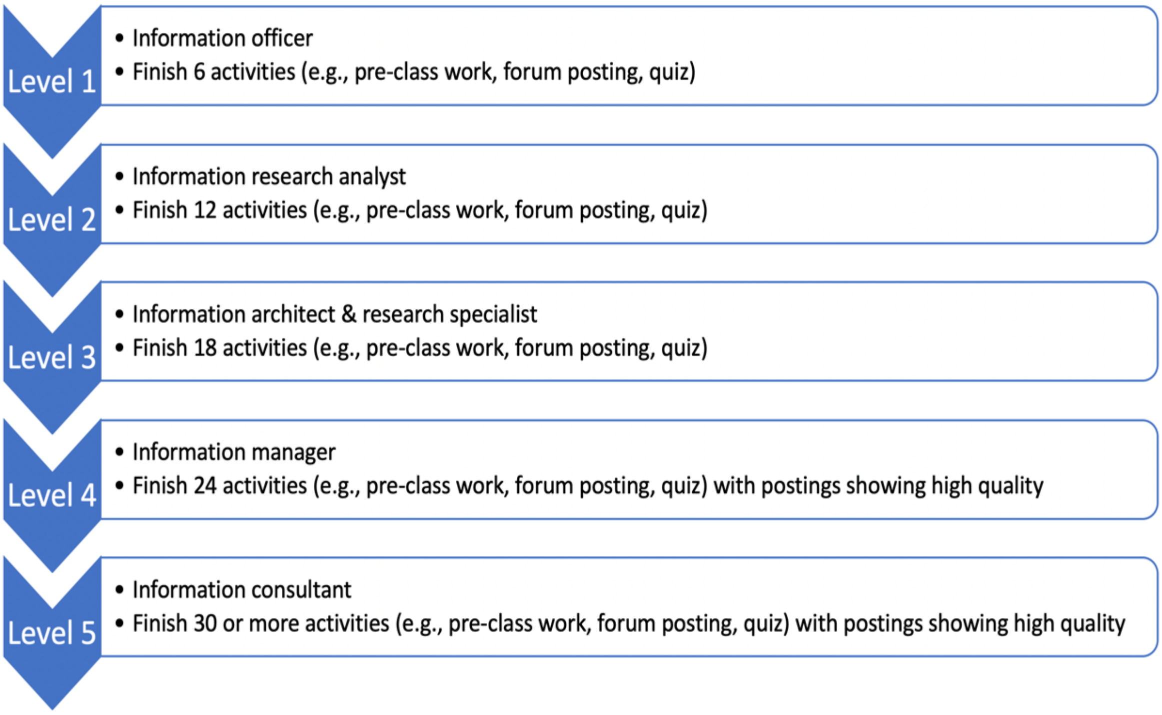


Figure 7. Game mechanics (levels) used in treatment group_1

However, a main concern was that some students frequently submitted low quality postings solely for the purpose of attaining participation badges. A number of students also perceived that the criteria for winning quality-based badges were not explicit enough. Although written explanations of how to win quality-based badges were provided in a slide on the course forum, few students clicked the slide and viewed the explanations. Based on the findings of pilot study 1, we determined that although gamification can help students set goals and provide feedback, the game mechanics may not sufficiently promote learning because students can trick the system by submitting low-quality content.

5. Main Study 2: Optimizing the use of gamification

5.1. Method

In response to the findings and concerns in pilot study 1, we proposed an updated-GAFCC model. We assumed that gamification design could be supplemented by other pedagogical strategies, such as teacher feedback, to optimize students' quality of submission. Thus, in the updated-GAFCC model, the five-step design process is supplemented with pedagogical strategies (Figure 8). We tested the effectiveness of the updated-GAFCC model with a new cohort (i.e., treatment group_2) and compared the results with those for treatment group_1 and the control group, as shown in Figure 9. Before beginning the experiment, we briefed the students about the research project, and invited them to sign a consent form. In total, 50 students participated in treatment group_2. Pre-test results showed that there were no significant differences in the students' prior knowledge between the three groups.

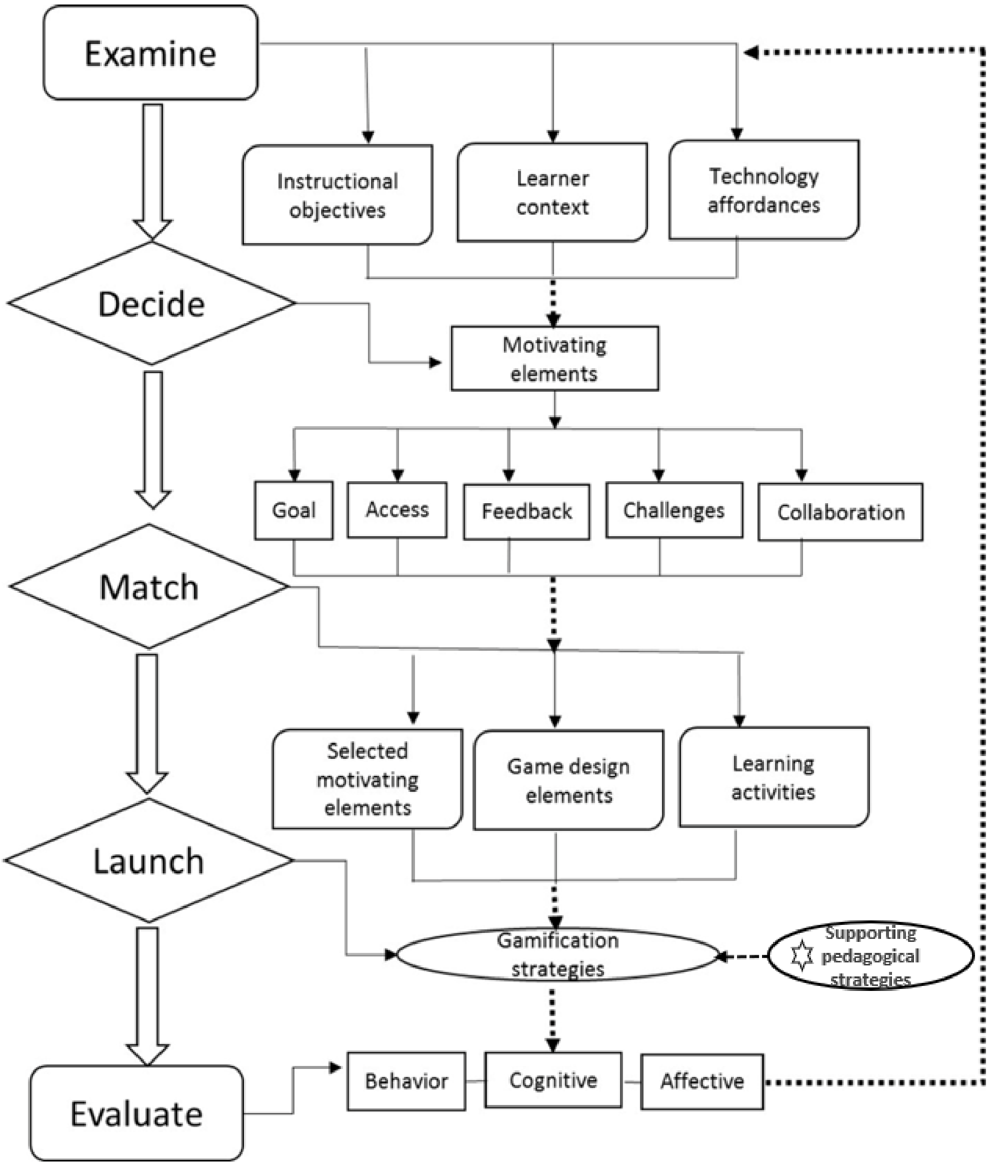
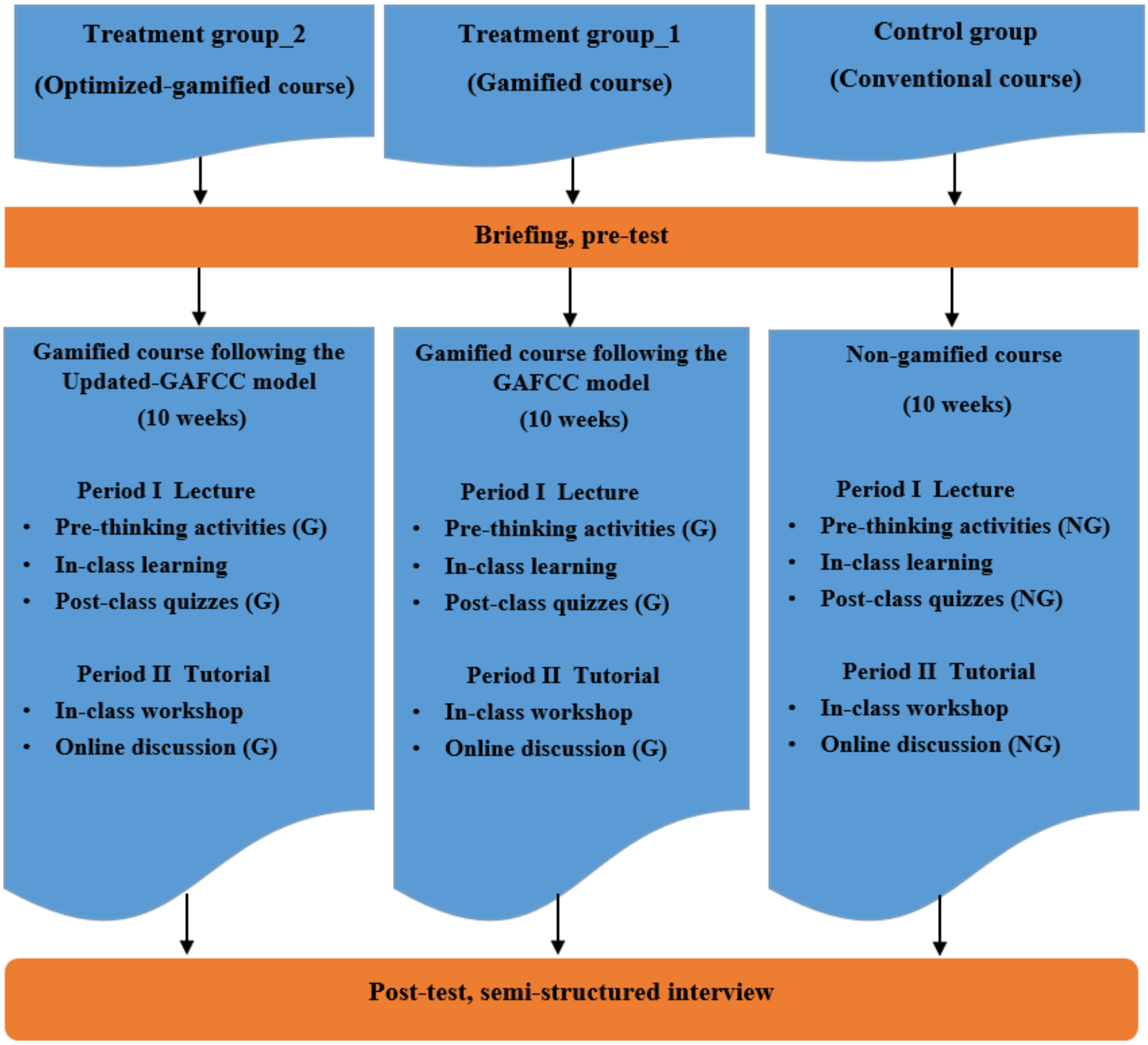


Figure 8. The five-step design process in the updated-GAFCC model



Note. "G" refers to "gamified," "NG" refers to "non-gamified."
Figure 9. Procedures of the design-based research

5.2. Instructional design of Study 2

The design process consisted of the following phases.

5.2.1. Phase 1: Analysis of the problems revealed in Study 1

In the analysis phase, the researcher collaborated with the course instructor to solve the problems that arose in Study 1. Although most of the students in Study 1 preferred to learn in a gamified environment, they reported a number of specific problems (Figure 10). One student stated that the timing of the vote for the best team was too late, and suggested holding it earlier. In addition to the problems raised by the students, the researcher and the course instructor were aware that insufficient feedback on student work was provided in Study 1, and the leaderboard display was not convenient to access.

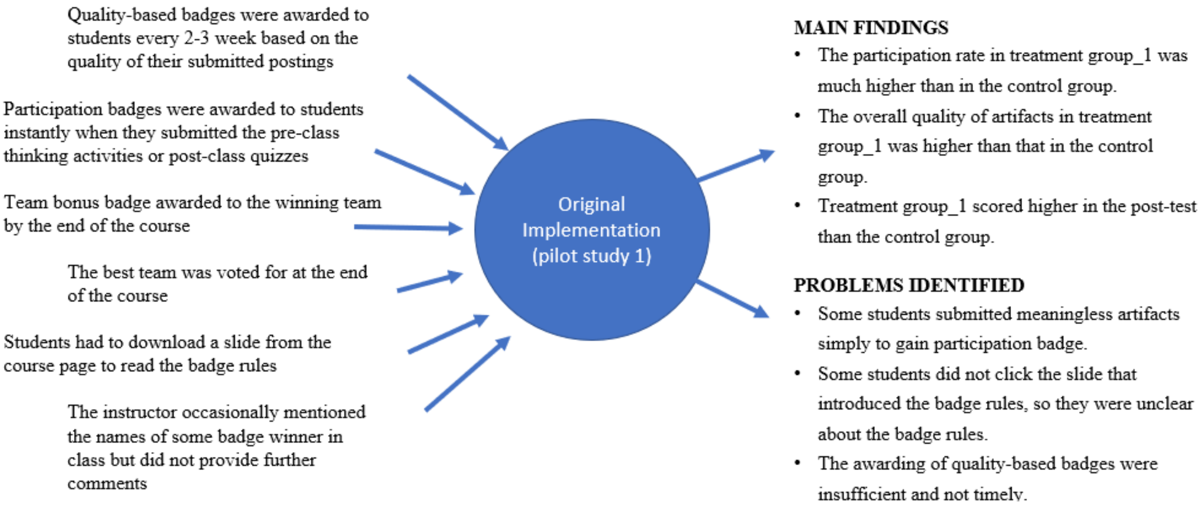


Figure 10. Summary of the implementation, findings, and problems in pilot study 1

5.2.2. Phase 2: Design and construction of the further-refined solutions in practice

The same intended outcomes and game elements (Figure 6) were used in the main study. The researchers reflected on the problems identified in the previous cycle, and proposed to refine the design by following the updated-GAFCC design model. With respect to the gamification strategies, the researchers decided to do the following:

- Provide a more direct way to visualize the badge rules and different levels. The badge rules and levels were displayed directly on the main announcement page, which could be immediately seen upon login, so the students did not need to download a slide to view the rules or levels (Figure 11).

- Assign badges and team bonuses more frequently. Quality-based badges were assigned to students on a weekly basis, rather than every two to three weeks, and team bonuses were assigned on a weekly basis rather than at the end of the course.

- Move voting for the best team to the middle of the course rather than at the end of the course.

With respect to the pedagogical strategies, it was decided that the instructor would provide elaborative feedback to students at the beginning of the class at least once every two weeks. Feedback is most beneficial when it guides the modification of thinking and provides students with opportunities for refinement (National Research Council [NRC], 2000). Although gamification can provide evaluative feedback regarding what learners have achieved, such as granular feedback (e.g., displaying points linked to each positive action), sustained feedback (e.g., using a performance graph to display progress), and cumulative feedback (e.g., tracking and assessing a series of learner actions and presenting the results on a leaderboard) (Rigby & Ryan, 2011), it does not convey complex information on what the learners can do to improve their thinking. For this reason, supplementing the gamified feedback with elaborative feedback from the teacher would likely help the students to better understand how well they have performed and how they can improve the quality of their submissions. For example, in the pre-class thinking activities, the students had made their thinking visible to themselves, teachers, and peers. Thus, if the students subsequently listened to the teacher's elaborative feedback, they would likely to reflect on their thinking and refine their submissions in future weeks (NRC, 2000). In the earlier study, the instructor only occasionally mentioned the names of the students who had won specific badges and did not make any further comments. In this cycle, the instructor provided examples of high-quality content based on the students' submissions and explained why the examples were considered to be of high quality. See Figure 12. for the procedure of the elaborative feedback.

5.2.3. Phase 3 Test of the proposed solutions

After updating the design, the course instructor implemented the refined gamification strategies and pedagogical strategies in treatment group_2. The efficacy of the design was evaluated by analyzing the students' learning artifacts, participation data, and interview results.

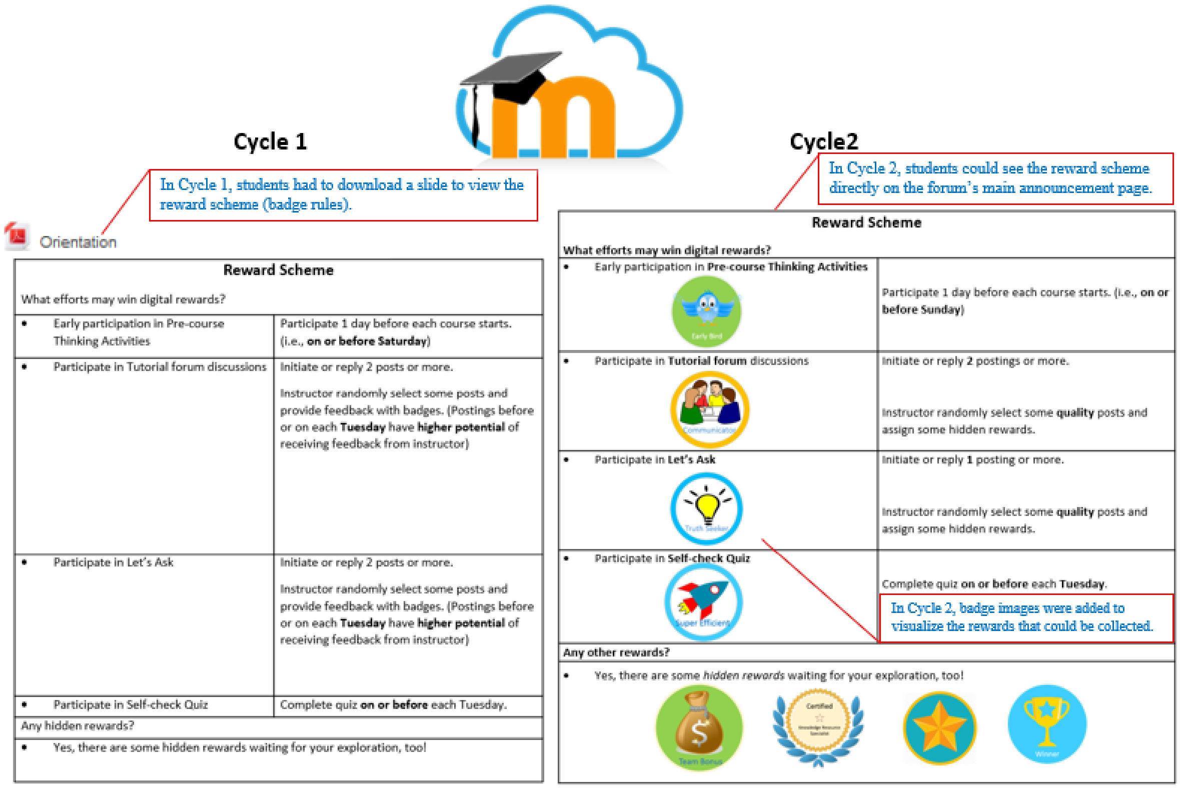


Figure 11. Display of badge rules on Moodle in Cycle 1 and Cycle 2

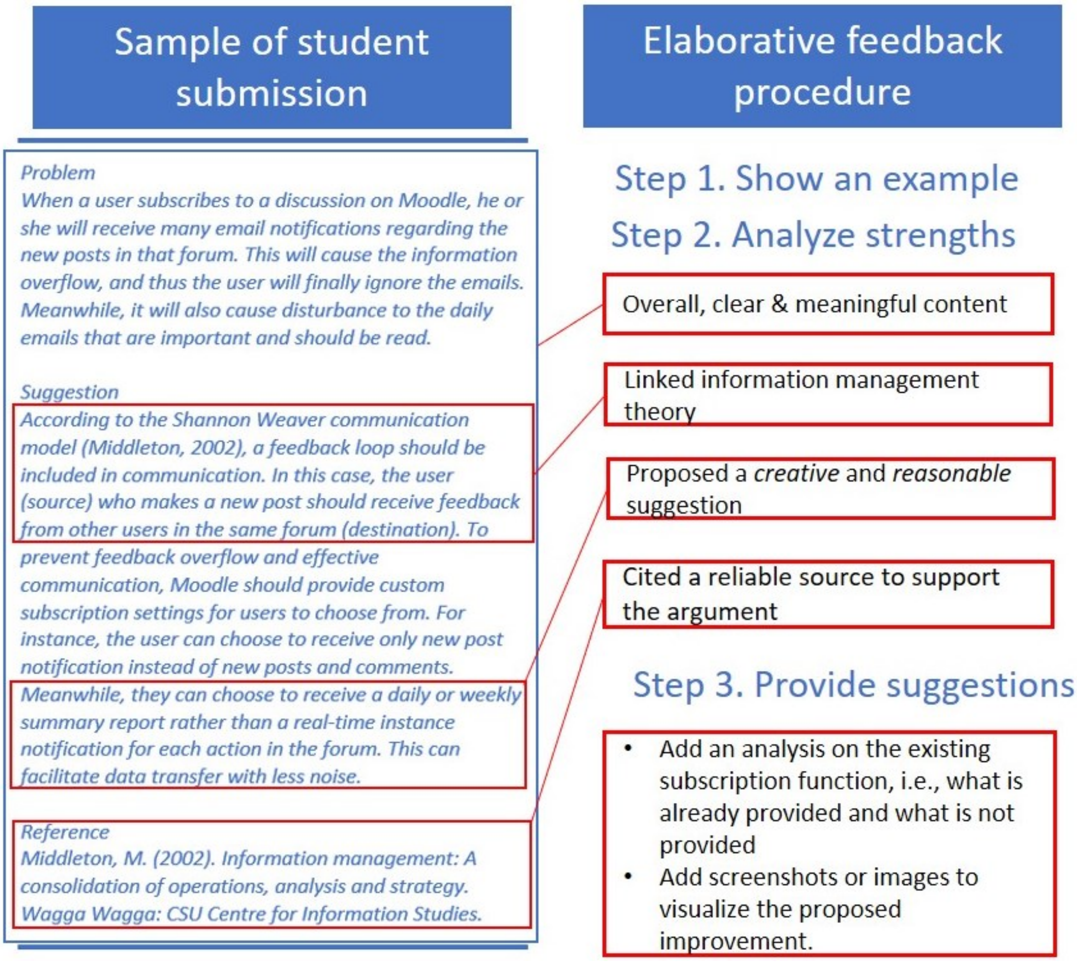


Figure 12. A sample of submissions and the procedure of elaborative feedback in Cycle 2

6. Results

To compare the effectiveness of the conventional approach, the GAFCC model, and the updated-GAFCC model, we evaluated the students' participation data and the quality of the artifacts. The participation indicators were the completion of pre-class thinking activities and post-class quizzes. The quality indicators were the pre- and post-test scores, levels of critical thinking exhibited in the thinking activities, and scores for the post-class quizzes.

6.1. Impact on participation

6.1.1. Pre-class activity completion

Pre-class thinking activities were assigned to the students each week, except for week 1 and week 10. In weeks 4 and 5, only one pre-thinking activity (i.e., pre-thinking activity 4-5) was assigned. A technical problem occurred in pre-thinking activity 4-5 in the first year, and the students could not submit their work for the activity. Hence, pre-thinking activity 4-5 was not included in the analysis. Chi-square tests of independence showed that there were significant differences in the completion rates of the three groups during four of the six weeks examined, i.e., weeks 6, 7, 8, and 9. Treatment group_2 and treatment group_1 were more likely to complete the activities before the due date than the control group. For example, a significant result was found in week 6 ($X^2(2) = 33.97$, $p < .001$), which showed that treatment group_2 (76%) and treatment group_1 (75%) were more likely to complete the activities before the due date than the control group (25%) (Figure 13 and Table 1).

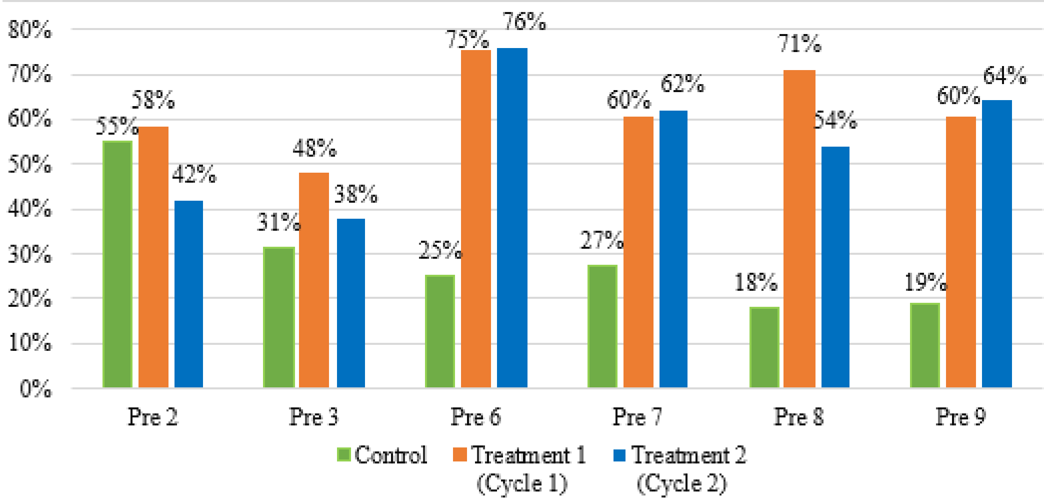


Figure 13. Pre-class thinking activities completed before the weekly due dates (Cycles 1 & 2)

Table 1. Completion of the pre-class thinking activities before the weekly due dates (Cycles 1 & 2)

Activity	Condition	Completed before the due date	Chi-square
Pre_think 2	Treatment2	42%	$X^2(2) = 2.84$, $p > .05$
	Treatment1	58%	
	Control	55%	
Pre_think 3	Treatment2	38%	$X^2(2) = 2.84$, $p > .05$
	Treatment1	48%	
	Control	31%	
Pre_think 6	Treatment2	76%	$X^2(2) = 33.97$, $p < .001$
	Treatment1	75%	
	Control	25%	
Pre_think 7	Treatment2	62%	$X^2(2) = 15.05$, $p < .05$
	Treatment1	60%	
	Control	28%	
Pre_think 8	Treatment2	54%	$X^2(2) = 27.20$, $p < .001$
	Treatment1	71%	
	Control	18%	
Pre_think 9	Treatment2	64%	$X^2(2) = 24.55$, $p < .001$
	Treatment1	60%	
	Control	19%	

Note. Treatment2 ($n = 50$); Treatment1 ($n = 48$); Control ($n = 48$).

6.1.2. Post-class quiz completion

Post-class quizzes were provided to the students each week, except for weeks 1 and 10. In weeks 4 and 5, only one quiz (i.e., quiz 4-5) was assigned to the students. Chi-square tests of independence showed significant differences in the completion rates of treatment group_2, treatment group_1, and the control group in five of the seven weeks, namely weeks 4-5, 6, 7, 8, and 9. Treatment group_2 and treatment group_1 were more likely to complete the activity before the due date than the control group. For example, a significant interaction was found in week 4-5 ($X^2(2) = 8.41$, $p < .05$), which showed that treatment group_2 (32%) and treatment group_1(25%) were more likely to complete the activity before the due date than the control group (15%) (see Figure 14 and Table 2).

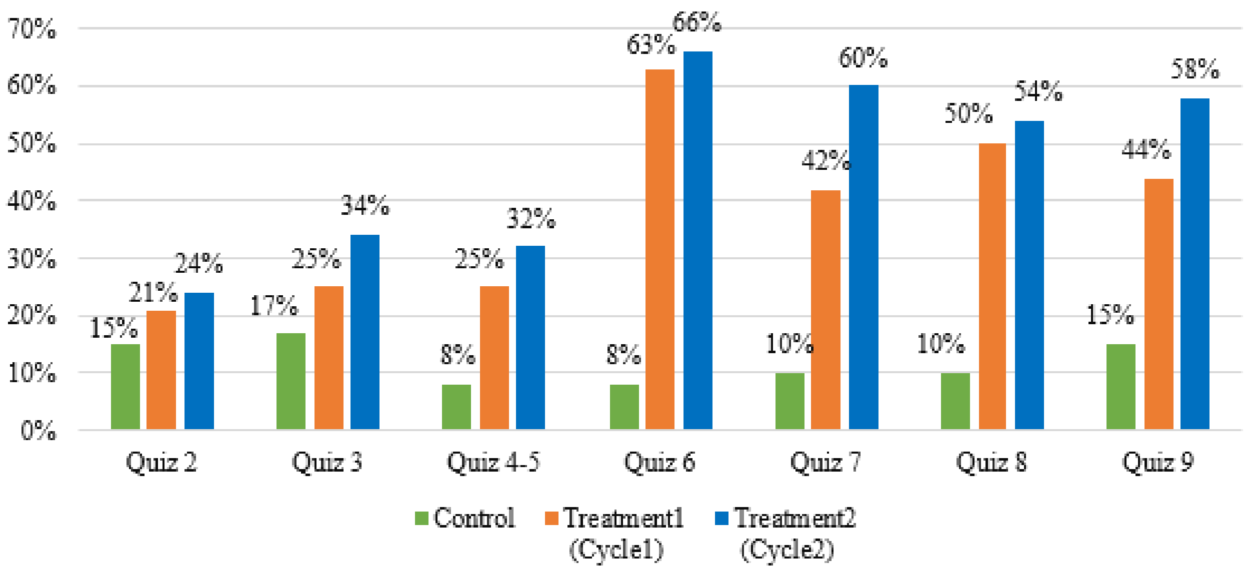


Figure 14. Post-class quizzes completed before the weekly deadlines (Cycles 1 & 2)

Table 2. Completion of post-class quizzes before the weekly due dates (Cycles 1 & 2)

Post-class quiz	Condition	Completed before the deadline	Chi-square
Quiz 2	Treatment2	24%	$X^2(2) = 1.41$, $p > .05$
	Treatment1	21%	
	Control	15%	
Quiz 3	Treatment2	34%	$X^2(2) = 3.89$, $p > .05$
	Treatment1	25%	
	Control	17%	
Quiz 4-5	Treatment2	32%	$X^2(2) = 8.41$, $p < .05$
	Treatment1	25%	
	Control	8%	
Quiz 6	Treatment2	66%	$X^2(2) = 40.74$, $p < .001$
	Treatment1	63%	
	Control	8%	
Quiz 7	Treatment2	60%	$X^2(2) = 26.13$, $p < .001$
	Treatment1	42%	
	Control	10%	
Quiz 8	Treatment2	54%	$X^2(2) = 23.78$, $p < .001$
	Treatment1	50%	
	Control	10%	
Quiz 9	Treatment2	58%	$X^2(2) = 20.06$, $p < .001$
	Treatment1	44%	
	Control	15%	

Note. Treatment2 ($n = 50$); Treatment1 ($n = 48$); Control ($n = 48$).

6.2. Impact on participation

6.2.1. Pre-test and post-test scores

As a normality test indicated that the pre-test scores were not evenly distributed, a nonparametric Kruskal-Wallis H test was used to examine the differences between the three groups. The Kruskal-Wallis H test revealed that there were no statistically significant differences in the pre-test scores of the different conditions, $X^2(2) = 2.27$, $p > .05$, with a mean rank score of 56.85 for treatment group_2 ($n = 40$), 65.21 for treatment group_1 ($n = 39$), and 58.08 for the control group ($n = 40$).

The post-test scores of the three groups were also evaluated. Kruskal-Wallis H test was used to examine the differences between the groups. The results showed that there were statistically significant differences in the post-test scores between the different conditions, $X^2(2) = 14.52$, $p < .001$, with a mean rank score of 56.66 for treatment group_2 ($n = 35$), 60.09 for treatment group_1 ($n = 38$), and 36.32 for the control group ($n = 30$). A post-hoc Bonferroni test showed there were significant differences between treatment group_2 ($M = 4.3$, $SD = 1.26$) and the control group ($M = 3.60$, $SD = 1.22$), $p = .008$. Similarly, there were significant differences between treatment group_1 ($M = 4.43$, $SD = 1.10$) and the control group ($M = 3.60$, $SD = 1.22$), $p = .001$. Both treatment group_2 and treatment group_1 outperformed the control group in the post-test.

6.2.2. Levels of critical thinking in pre-class thinking activities

The quality of the pre-class thinking artifacts was evaluated based on the modified Taxonomy of Critical Thinking (Greenlaw & Deloach, 2003). Greenlaw and Deloach (2003) proposed that the level of critical thinking can be measured with respect to the degree of sophistication in an argument. We adapted the taxonomy to evaluate thinking artifacts (as shown in Figure 15) by retaining level 0 (off-the-topic); combining the original level 1(unilateral description) and 2 (simplistic argument) into one level as both deal with simple description or justification; retaining level 3 (explicit analysis); and merging levels 4 (theoretical inference) and 5 (empirical inference) into one level, namely theoretical and/or empirical inference. Level 6, merging value with analysis, was not included as it was beyond the scope of our investigation. However, we included creativity as an additional factor, because it was frequently observed in the artifacts of treatment group_2.

Level 0	**Level 1-Lower level**
❑**Off-the-topic, no submission, or submission after the due date**	❑**Mere repetition or simplistic arguments** *e.g., repeating question statements without adding new information or interpretation* *e.g., making confusing or ambiguous statements* *e.g., assertions without evidence, or giving an example to provide a simple explanation*
Level 2-Upper level	**Level 3-Uppermost level**
❑**Serious attempts to analyze an argument or list competing arguments with evidence** *e.g., serious argument, such as listing factors as evidence, exploring competing argument, citing anecdotal evidence, but with logical flaws* *e.g., serious argument with at least 1 creative perspective (which can shed light on the issue/or can bring inspiration to other students) but with logical flaws* *e.g., serious argument but not referring to theory or empirical data*	❑**Serious argument with a clear logical framework, strengthened by empirical evidence and/or theory** *e.g., use historical data, citing reliable resources, using theory to test the validity of an argument* *e.g., serious argument with two or more creative perspectives (which can shed light on the issue or bring inspiration to other students) and a clear logical framework*

Figure 15. Revised Taxonomy of Critical Thinking (adapted from Greenlaw and Deloach, 2003)

In the data analysis, a researcher first coded all of the submissions. An independent observer then randomly selected and coded 20% of the submissions. The Krippendorff's alpha test, a widely adopted test for examining the inter-rater reliability of content analysis results (Krippendorff, 2004), was conducted. The Krippendorff's alpha test result was $\alpha = 0.88$, indicating a substantial level of agreement between the two coders.

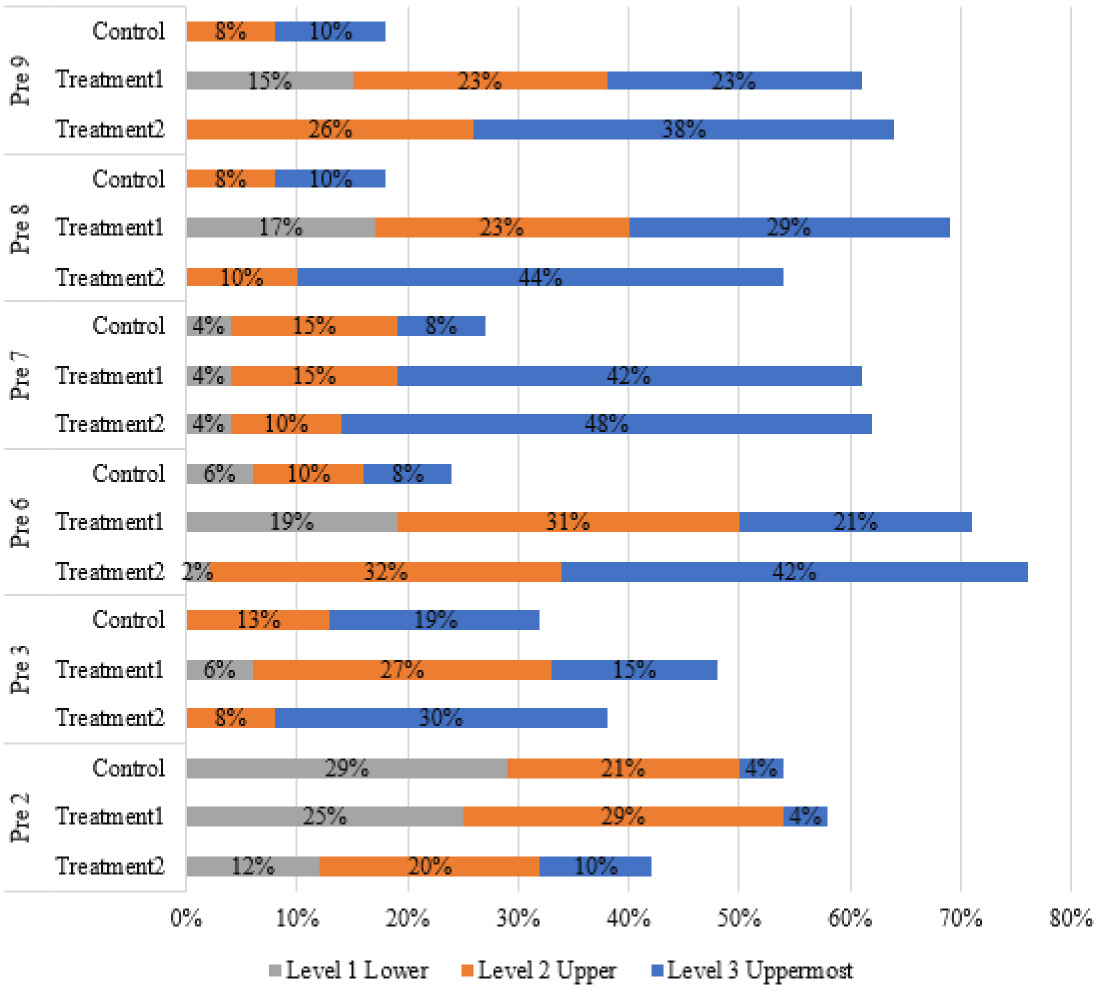


Figure 16. Distribution of the critical thinking levels in the pre-class thinking activities (Cycles 1 & 2)

Table 3. Quality of the pre-class thinking activities (Cycle 1 & 2)

Activity	Condition	Level 3 (uppermost)	Level 2 (upper)	Level 1 (lower)	Level 0 (low-late-no)	*p*-value
Pre_think 2	Treatment2	10%	20%	12%	58%	.246
	Treatment1	4%	29%	25%	42%	A, B = .134
	Control	4%	21%	29%	46%	A, C = .141
Pre_think 3	Treatment2	30%	8%	0%	62%	.019
	Treatment1	15%	27%	6%	52%	A, B = .008**
	Control	19%	13%	0%	69%	A, C = .432
Pre_think 6	Treatment2	42%	32%	2%	24%	< .001
	Treatment1	21%	31%	19%	29%	A, B = .013**
	Control	8%	10%	6%	75%	A, C < .001**
Pre_think 7	Treatment2	48%	10%	4%	38%	<.001
	Treatment1	42%	15%	4%	40%	A, B =.89
	Control	8%	15%	4%	73%	A, C < .001**
Pre_think 8	Treatment2	44%	10%	0%	46%	<.001
	Treatment1	29%	23%	17%	31%	A, B =.002**
	Control	10%	8%	0%	81%	A, C < .001**
Pre_think 9	Treatment2	38%	26%	0%	36%	<.001
	Treatment1	23%	23%	15%	40%	A, B = .020*
	Control	10%	8%	0%	81%	A, C < .001**

Note. *significant using $p < .05$; **significant using an adjusted-alpha of 0.017 (Bonferroni adjustment). Treatment2 ($n = 50$); Treatment1 ($n = 48$); Control ($n = 48$). A represents *treatment2*, B represents *treatment1*, and C represents *the control group*.

The descriptive data showed that treatment group_2 had more uppermost level submissions than treatment group_1 and the control group in all weeks. On average, 35% of treatment group_2, 22% of treatment group_1, and 10% of the control group students submitted uppermost level artifacts (Figure 16). As the expected counts of several cells were smaller than five, we conducted Fisher's exact tests rather than Chi-square tests of independence to compare the differences between the three groups since the Fisher's exact test is more appropriate for small values (Bolboacă et al., 2011). As shown in Table 3, the results indicated there were significant differences in the quality of the pre-class thinking artifacts of treatment group_2, treatment group_1, and the control group in five out of six weeks, namely weeks 3, 6, 7, 8, and 9. For example, in week 6, the students in treatment group_2 (42%) were more likely to submit uppermost level artifacts than those in treatment group_1 (21%) and the control group (8%), $p < .001$. Pairwise Fisher's exact tests showed that treatment group_2 submitted significantly more uppermost level artifacts than treatment group_1 in weeks 3, 6, 8, and 9, and than the control group in weeks 6, 7, 8, and 9.

6.2.3. Performance in post-class quizzes

The quality of the post-class quizzes completed before the due dates for treatment group_2 was compared with that for treatment group_1 and the control group. The quizzes submitted after the due dates and absent submissions were recorded as 0. Kruskal-Wallis H tests results indicated that there were significant differences in the post-class quiz scores of the three groups, except in weeks 2 and 3 (Table 4). For example, in week 4-5, the results showed that there were statistically significant differences in the post-class quiz scores of the three conditions, $X^2(2) = 9.67$, $p = .008$, with a mean rank score of 81.83 for treatment group_2 (n = 50), 75.38 for treatment group_1 (n = 48), and 62.95 for the control group (n = 48). Post-hoc Bonferroni tests showed that treatment group_2 outperformed the control group in weeks 4-5, 6, 8, and 9, and treatment group_1 outperformed the control group in weeks 6, 8, and 9. No significant differences were observed between treatment group_2 and treatment group_1, with the two groups performing equally well in the post-class quizzes.

Table 4. Post-class quiz scores completed before the due dates (Cycles 1 and 2)

	Group	*n*	Mean rank	Kruskal-Wallis test
Quiz 2	Treatment2	50	74.62	$X^2(2) = 0.848$, $p = .654$
	Treatment1	48	75.41	
	Control	48	70.43	
Quiz 3	Treatment2	50	80.5	$X^2(2) = 4.802$, $p = .091$
	Treatment1	48	73.38	
	Control	48	66.33	
Quiz 4-5	Treatment2	50	81.83	$X^2(2) = 9.67$, $p = .008$
	Treatment1	48	75.38	A, B = .88
	Control	48	62.95	A, C = .010**
Quiz 6	Treatment2	50	88.79	$X^2(2) = 36.15$, $p < .001$
	Treatment1	48	84.32	A, B = 1
	Control	48	46.75	A, C < .001**
Quiz 7	Treatment2	50	89.5	$X^2(2) = 24.72$, $p < .001$
	Treatment1	48	76.69	A, B = 0.24
	Control	48	53.65	A, C < .001**
Quiz 8	Treatment2	50	81.53	$X^2(2) = 20.10$, $p <.001$
	Treatment1	48	84.07	A, B = 1
	Control	48	54.56	A, C = .001**
Quiz 9	Treatment2	50	88.17	$X^2(2) = 20.83$, $p < .001$
	Treatment1	48	76.8	A, B = .37
	Control	48	54.92	A, C < .001**

Note. **significant using an adjusted-alpha of 0.017 (Bonferroni adjustment). Treatment2 (n = 50); Treatment1 (n = 48); Control (n = 48). A represents *treatment2*, B represents *treatment1*, and C represents *the control group*.

6.3. Students' perceptions of gamification

The students in treatment group_2 were invited to participate in semi-structured interviews after the completion of the course. Nine students (five females and four males) volunteered to participate in the interviews. Sample interview questions were "What is your general feeling about the course?" "What are the positive/negative aspects of gamification?" "Do you prefer to study in a gamified or non-gamified environment? Why?" "Do you have any suggestions for improving the gamification design?"

The interview results were coded by two researchers. The first researcher read through all the transcripts and determined the coding scheme, including the themes, initial codes, and coding procedures. Examples of each theme were provided to increase the consistency of classification. Then, the researcher coded the transcripts. The constant-comparative approach (Glaser, 1965) was adopted, and any new categories that emerged were given a new label. After receiving training, the second researcher randomly selected 20% of the transcripts and coded them. The percent agreement (Campbell et al., 2013) between the two raters was 83%. Discrepancies were discussed until agreements reached.

The responses to the question, "Do you prefer to study in a gamified or non-gamified environment" indicated that 89% (eight out of nine) of the students preferred to learn in a gamified learning environment. The remaining student stated that she would have preferred to learn in a gamified learning environment had the leaderboard been more competitive. As to the question of the positive aspects of gamification, students commented that gamification encouraged peer collaboration and interaction (100%, *n* = 9), provided recognition (100%, *n* = 9), added an element of fun to the course (78%, *n* = 7), and enhanced their competence (67%, *n* = 6).

The students explained how gamification motivated them to collaborate more. For example, one student commented:
I thought the forum activities were a good way for us to communicate with other students. Even though we did not meet face to face (after class), we could still communicate wherever and whenever. These badges brought us closer. Because we had to talk about tasks and communicate constantly. (S7)

The recognition conveyed through gamification stimulated the students to treat the activities seriously. One student stated:
Because it felt like we were winning a reward or extra marks. Moreover, some of the hidden rewards were sent by e-mail, and after I received them, I was surprised and felt that I had done a great job in the forum. This encouraged us to do more and take the tasks seriously. (S4)

The students also explained why it is important to include fun and enjoyment in a course. For instance, one student stated:
I enjoyed the gamified course because it was fun. I am a person who believes that if you enjoy learning, you will learn more than you would otherwise. People could enjoy this course, and I think that is essential. (S5)

In the updated-GAFCC condition, students focused more on making quality submissions rather than the quantity of their submissions (78%, *n* = 7). For example, one student expressed, "I put a lot more effort into this course than other courses. I spent a lot of time reading the notes so as to submit quality posts." (S1) Another student expressed that, "as gamification was used in this course, I decided that I not only had to use my own knowledge to reply on the forum, but also had to provide a quality response. It encouraged me to learn more about the topic. For example, I did more reading and things like that. It helped us discover different ways to improve our knowledge of this topic." (S7) Students also admired the quality submissions of their peers, and perceived reading these submissions as a learning opportunity. One student stated, "I appreciate that some of them have done a really great job. The posts were long and full of very meaningful content. I could learn more stuff by reading their posts." (S4)

The rise in eagerness to submit quality posts could be attributed to the adjustment of gamification strategies, such as the more frequently assigned quality-based badges and the display of badge rules directly on the main announcement page. These adjustments helped to emphasize the expected performance goals in this course.

The supporting pedagogical strategies (i.e., instructor elaborative feedback) empowered students to make quality submissions. One stated that, "the instructor gave us a lot of constructive feedback throughout the course, so it really helped me." (S6) She also commented that the learning design not only encouraged them to think about the completion of the activity that time, but also how they could do better next time.

Furthermore, the students made a number of suggestions, including rewarding students with real gifts or coupons (44%, *n* = 4) and reducing the workload by combining activities or removing tasks during the midterm exams (44%, *n* = 4). Two thirds of the students reported that their favorite game mechanics were the quality-based badges (66%, *n* = 6), such as the coffee coupons and tour package (33%, *n* = 3), because they were challenging to win, and the team bonus (33%, *n* = 3), which brought their teams closer together and made them feel rewarded. Similar to treatment group_1, six of the nine students mentioned that the sense of closeness between their teammates had been strengthened. Most importantly, 89% (eight out of nine) of the students thought that the overall learning design, i.e., the integration of gamification strategies and learning activities, was quite successful. This figure was much higher than that for treatment group_1 (50%). Overall, the students felt that the design of the learning tasks and the use of gamification strategies worked well together and ultimately facilitated their learning.

7. Discussion

The aim of this study is to test and refine a theory-based gamification design model, and examine its effects on students' learning achievement, participation, and perceptions. Specifically, we refined the GAFCC design model and applied it to three cohorts of students in an undergraduate course. The results indicated that the updated-GAFCC model were effective in enhancing learning achievements and participation, as well as motivating more students to produce uppermost level thinking artifacts. The findings contribute to our understanding of how the updated theory-based design model can be used to optimize learning outcomes.

7.1. Impact on learning achievements

In the updated-GAFCC condition, the quality of the thinking artifacts was higher than that in the GAFCC condition and the control condition. In the updated-GAFCC condition, many students used empirical data or theories to support their arguments within a clear logical framework. At the same time, creative arguments that could inspire others were frequently observed in the updated-GAFCC condition. In the GAFCC condition, more high-quality artifacts were produced than in the control group, but some students submitted low-quality artifacts (ranging from 4%-25%). This situation was much improved in the updated-GAFCC condition, with the proportion of lower quality artifacts reducing to 0% to 12%. The positive change confirmed the effectiveness of the updated-GAFCC model in promoting higher quality submissions. Research has suggested that more detailed feedback can guide the modification of students' thinking and help them understand what constitutes a quality submission (NRC, 2000). Thus, the students in the updated condition had a clearer understanding of the quality goals they needed to achieve. This result is consistent with the suggestion that gamified feedback and teacher elaborative feedback can be used together to optimize the learning outcomes in a gamified environment (Lo & Hew, 2018). At the same time, our adjustment of the gamification design, in which we presented the game rules in a more obvious section of the course page, may have raised the students' awareness of the gamified activities and promoted the internalization of certain behavioral and cognitive goals. The more timely quality-based badges also provided more immediate recognition and confirmation of students' positive behavior. Together, the adjustment of the gamification strategies and the inclusion of a supporting pedagogy motivated the students to make more in-depth submissions. In addition, more creative submissions were observed in the updated-GAFCC condition, which is consistent with the finding of the interviews that the students practiced thinking beyond the questions given by the lecturer.

The updated-GAFCC condition and the GAFCC condition performed equally well in the post-tests and post-quizzes. An interesting observation was that when completing the less challenging tasks in the post-class quizzes, such as recalling factual knowledge, the students' submissions in the two gamified conditions were of similar quality. However, when completing more complex tasks, such as in the thinking activities, the gap between the two conditions increased, with the updated-GAFCC condition stimulating more higher level critical thinking than the GAFCC condition.

7.2. Impact on task completion

More students in the updated-GAFCC condition and the GAFCC condition completed the pre-thinking activities before the due date than in the control group. The average completion rates before the due date were 29% for the control group, 62% for the gamified group, and 56% for the updated-GAFCC condition. Moreover, significantly more students in the updated-GAFCC condition and the GAFCC condition completed the post-class quizzes than in the non-gamified group in five of the seven weeks. Students in the updated-GAFCC condition explained that they felt that the content of the gamified activities was designed to enhance their knowledge and skills, and that the use of gamification strategies, such as badges and leaderboards, provided a source of positive encouragement that motivated them to participate in the activities. Moreover, the students felt they had a choice regarding whether to participate in the activities, and they did not feel a sense of being manipulated. This provided the students with a sense of control over their own learning (Deci et al., 1981) and facilitated motivational learning experiences.

7.3. Impact on perception

The interview results indicated that in the updated-GAFCC condition, 89% of the students preferred to study in a gamified environment, and that 11% would opt to learn in a gamified environment if a more competitive leaderboard was introduced. In the updated-gamified condition, the majority of the students preferred to study in a gamified environment because they could collaborate and interact more with their peers and found the experience enjoyable (Zatarain-Cabada et al., 2020). The students thought their efforts were recognized in the gamified condition (Mullins & Sabherwal, 2020), and it enhanced their sense of achievement. The high-quality thinking submissions also provided them with opportunities to learn from their peers. In sum, the students were satisfied with the overall learning design.

8. Conclusion

8.1. Impact on learning achievements

Our three-year design-based study was carried out in collaboration with the course instructor. In addition, feedback was collected from students to inform the design. This study demonstrated how the updated-GAFCC design model can be implemented in courses to engage students on the cognitive and behavioral levels. Several lessons can be learned from the design process implemented in this study: (1) it is important to build alignment between the instructional objectives, game rules, and game elements; (2) the students' sense of feeling instrumental is key to the success of a gamification design; (3) the use of gamification can enhance student learning and task completion; (4) the game rules should be displayed in a prominent place in the classroom or on the website for students to view; (5) timely elaborative feedback helps build competence; (6) lastly, iterative design and refinement are necessary for developing mature learning designs.

8.2. Limitations

This study has a number of limitations that should be noted when interpreting the results. First, we tested the updated-GAFCC design model with students enrolled in an information management course. The nature of this course may differ from that of other disciplines, such as computer programming and business education. Thus, further research is warranted on the impact of the updated-GAFCC design model in other disciplines. Second, the sample used in this study was only representative of Asian students in a higher education setting, and the learner characteristics and user types may have influenced their acceptance of the gamification strategies and the overall learning design. Moreover, it should be noted that the interviews were conducted on a voluntary basis. Therefore, the students who participated in the interviews were more likely to have been those who had higher motivation levels (Robinson, 2014), and so we have to be cautious with regards to the generalizability of the collected information. In the future, it would be exciting to apply the updated design model in other cultures and compare the results across settings.

Acknowledgement

The research was supported by a grant from the Research Grants Council of Hong Kong (Project reference no: 17616020).

References

Aguilar, S. J., Holman, C., & Fishman, B. J. (2018). Game-inspired design: Empirical evidence in support of gameful learning environments. *Games and Culture*, *13*(1). 44-70.

Amiel, T., & Reeves, T. C. (2008). Design-based research and educational technology: Rethinking technology and the research agenda. *Journal of Educational Technology and Society*, *11*(4), 29-40.

Bai, S., Hew, K. F., & Huang, B. (2020). Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review*, *30*, doi:10.1016/j.edurev.2020.100322

Barata, G., Gama, S., Jorge, J., & Goncalves, D. (2017). Studying student differentiation in gamified education: A Long-term study. *Computers in Human Behavior, 71*, 550-585.

Baydas, O., & Cicek, M. (2019). The Examination of the gamification process in undergraduate education: A Scale development study. *Technology, Pedagogy and Education*, *28*(3), 269-285.

Bolboacă, S. D., Jäntschi, L., Sestras, A. F., Sestras, R. E., & Pamfil, D. C. (2011). Pearson-Fisher Chi-Square statistic revisited. *Information*, *2*(3), 528-545.

Cai, F., Dai, G., & Han, T. (2016). Gamification design based research on fitness mobile application for university students. In A. Marcus (Ed.), *Design, user experience, and usability: Novel user experiences*. *(pp.240-251).* Cham, Switzerland: Springer.

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding In-depth Semistructured Interviews: Problems of Unitization and Intercoder Reliability and Agreement. *Sociological Methods & Research*, *42*(3), 294–320.

Csikszentmihalyi, M. (1990). *Flow: The Psychology of optimal experience*. New York, NY: Harper & Row.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227-268.

Deci, E. L., Nezlek, J., & Sheinman, L. (1981). Characteristics of the rewarder and intrinsic motivation of the rewardee. *Journal of Personality and Social Psychology*, *40*(1), 1-10.

Deterding, S., Khaled, R., Nacke, L. E. & Dixon, D., (2011). From Game Design Elements to Gamefulness: Defining "Gamification". *In Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments* (pp.9-15). New York, NY:: Association for Computing Machinery.

Dichev, C., & Dicheva, D. (2017). Gamifying education: What is known, what is believed and what remains uncertain: A Critical review. *International Journal of Educational Technology in Higher Education, 14*(9), doi:10.1186/s41239-017-0042-5

Ding, L. (2019). Applying gamifications to asynchronous online discussions: A Mixed methods study. *Computers in Human Behavior*, *91*, 1-11.

Festinger, L. (1954). A Theory of social comparison processes. *Human Relations, 7*(2), 117-140.

Glaser, B. G. (1965). The Constant comparative method of qualitative analysis. *Social Problems*, *12*(4), 436–445.

Greenlaw, S. A., & DeLoach, S. B. (2003). Teaching critical thinking with electronic discussion. *The Journal of Economic Education, 34*(1), 36-52.

Huang, B., & Hew, K. F. (2018). Implementing a theory-driven gamification model in higher education flipped courses: Effects on out-of-class activity completion and quality of artifacts. *Computers & Education*, *125*, 254-272.

Huang, B., Hew, K. F., & Lo, C. K. (2019). Investigating the effects of gamification-enhanced flipped learning on undergraduate students' behavioral and cognitive engagement. *Interactive Learning Environments*, *27*(8), 1106-1126.

Jo, J., Jun, H., & Lim, H. (2018). A Comparative study on gamification of the flipped classroom in engineering education to enhance the effects of learning. *Computer Applications in Engineering Education*, *26*(5), 1626-1640. doi:10.1002/cae.21992

Jong, M. S. Y. (2019). Sustaining the adoption of gamified outdoor social enquiry learning in high schools through addressing teachers' emerging concerns: A 3-year study. *British Journal of Educational Technology*, *50*(3), 1275-1293.

Krippendorff, K. (2004). *Content analysis: An Introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage.

Kyewski, E., & Kramer, N. C. (2018). To gamify or not to gamify? An Experimental field study of the influence of badges on motivation, activity, and performance in an online learning course. *Computers & Education, 118*, 25-37.

Landers, R. N., Bauer, K. N., & Callan, R. C. (2017). Gamification of task performance with leaderboards: A Goal setting experiment. *Computers in Human Behavior*, *71*, 508-515.

Lee, J. J., Matamoros, E., Kern, R., Marks, J., de Luna, C., & Jordan-Cooley, W. (2013). Greenify: Fostering sustainable communities via gamification. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (pp. 1497-1502). New York, NY: Association for Computing Machinery.

Li, W., Grossman, T., & Fitzmaurice, G. (2012). GamiCAD: A Gamified tutorial system for first time Autocad users. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology* (pp. 103-112). New York, NY: Association for Computing Machinery.

Lo, C. K., & Hew, K. F. (2018). A Comparison of flipped learning with gamification, traditional learning, and online independent study: The effects on students' mathematics achievement and cognitive engagement. *Interactive Learning Environments*, *28*(4), 464-481.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, *57*(9), 705-717.

Lopez, C. E., & Tucker, C. S. (2019). The Effects of player type on performance: A Gamification case study. *Computers in Human Behavior*, *91*, 333-345.

McKenney, S., & Reeves, T. (2012). *Conducting educational design research*. New York, NY: Routledge.

Mullins, J. K., & Sabherwal, R. (2020). Gamification: A Cognitive-emotional view. *Journal of Business Research, 106*, 304-314.

National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. Washington, DC: The National Academies Press. doi:10.17226/9853

Rigby, C. S., & Ryan, R. M. (2011). *Glued to games: How video games draw us in and hold us spellbound*. Santa Barbara, CA: Praeger.

Robinson, O. C. (2014). Sampling in interview-based qualitative research: A Theoretical and practical guide. *Qualitative Research in Psychology*, *11*(1), 25-41.

Sailer, M., & Sailer, M. (2020). Gamification of in-class activities in flipped classroom lectures. *British Journal of Educational Technology*. doi:10.1111/bjet.12948

Skinner, B. F. (1953). *Science and human behavior*. New York, NY: Macmillan.

Yildirim, I. (2017). The Effects of gamification-based teaching practices on student achievement and students' attitudes toward lessons. *The Internet and Higher Education*, *33*, 86-92.

Zainuddin, Z., Chu, S. K. W., Shujahat, M., & Perera, C. J. (2020). The Impact of gamification on learning and instruction: A Systematic review of empirical evidence. *Educational Research Review*, *30*, doi:10.1016/j.edurev.2020.100326

Zatarain-Cabada, R., Barrón Estrada, M. L., Ríos Félix, J. M., & Alor Hernández, G. (2020). A Virtual environment for learning computer coding using gamification and emotion recognition. *Interactive Learning Environments*, *28*(8), 1048-1063.

Organizing and Hosting Virtual PPTELL 2020 During the COVID-19 Pandemic

Yu-Ju Lan[1*], Kao Chia-Ling Gupta[2], Tai-Yi Huang[3], Shobhana Chelliah[3] and J. Michael Spector[3]

[1]National Taiwan Normal University, Taiwan // [2]University of Hong Kong, China // [3]University of North Texas, USA // yujulan@gmail.com // kclgupta@gmail.com // Tai-yi.huang@unt.edu // Shobhana.Chelliah@unt.edu // Mike.Spector@unt.edu

[*]Corresponding author

(Submitted September 10, 2020; Revised October 18, 2020; Accepted November 6, 2020)

ABSTRACT: This paper aims at answering the "how" questions about organizing and hosting an online conference during the COVID-19 pandemic. The 3rd International Pan-Pacific Technology-Enhanced Language Learning (PPTELL) Conference and Critical Thinking Meeting (hereafter PPTELL 2020) hosted from June 29 to July 1, 2020, on Zoom is the example conference used in this paper to illustrate the challenges and approaches adopted before, during and after the conference. The mentioned conference was supposed to take place physically at the University of North Texas during the same period but was transformed into an online virtual conference due to the outbreak of COVID-19 in early 2020. It was an urgent decision, along with many unknown situations, such as the attendees' different time zones and "Zoombombing." A three-staged and target-action process guided the preparation and organization of the online conference, i.e., pre-, during, and post-conference. According to the live meeting results and the post-conference survey, PPTELL 2020 has earned a reputation from its quality and the satisfaction of the participants and attendees. Therefore, the experience shared in the current paper seems to be a good reference for the organizers and hosts of international conferences.

Keywords: Online conference, COVID-19, Zoombombing, PPTELL 2020 and Critical Thinking Meeting

1. Introduction

1.1. Background

COVID-19 seems a "bomb" to the globe this year. It has been changing and influencing the world since early 2020, starting from Asia, Europe, Pan Pacific, and then America, and finally, it has become a global pandemic (Cucinotta & Vanelli, 2020). The outbreak has affected busy and frequent international communication (Verbeke, 2020). Globalization in the recent decade has fostered the spread of COVID-19. To reduce its damage and spread, social distancing is one of the most important measures. Consequently, imposing a shutdown seems to be a quick and effective approach to dealing with the plague (Cullinane & Montacute, 2020). Starting with Wuhan, China, several big cities in Europe, and then almost all the States in the U.S. declared a shutdown. The world seems to have come to a halt.

Following the shutdown and social distancing in many areas worldwide, many activities are hugely affected. In addition to the global economy (McKibbin & Fernando, 2020), education and academia (Crawford et al., 2020) are also among the most affected fields. According to the report of United Nations Educational Scientific and Cultural Organization (UNESCO) (2020), "the COVID-19 pandemic has globally affected the educational system, leading to the near-total closures of schools, universities, and colleges." The closures have impacted over 60% of the student population worldwide, reaching 1.725 billion students. To help continue students' learning, remote instruction, and Massive Open Online Courses (MOOCs) and Open Educational Resources (OER) (Miró, Baquero-Arnal, Civera, Turró, & Juan, 2018) have become the potential approaches to providing students with learning opportunities during the pandemic (Hartshorne, Baumgartner, Kaplan-Rakowski, Mouza, & Ferdig, 2020).

Similar to the closure of the educational system, holding academic activities, especially international conferences, requires new approaches to dealing with the chaos caused by the COVID-19 outbreak. Interpersonal interaction is the main features of such academic activities, but it is strongly prohibited for slowing the spread of the plague. In the fact of the pandemic, many international conferences in 2020 have been cancelled (Klöwer, Hopkins, Allen & Higham, 2020). Some are rescheduled (e.g., CALICO 2020 and CALL 2020) and transformed (e.g., from Teaching Professor Conference 2020 to Teaching Professor Virtual Conference).

The 3rd Pan-Pacific Technology-Enhanced Language Learning International Conference and Critical Thinking Meeting (hereafter PPTELL2020) faced the same dilemma as many other international conferences mentioned above. PPTELL was first organized in 2018 to respond to the rising trends of modern technologies for competency-oriented language learning. PPTELL Conference connects researchers, educators, and the front-line teachers to have conversations on the huge potential of integrating language learning theories and advanced technologies for cultivating learners' critical competencies for pursuing success in the 21st century. PPTELL 2020 was co-hosted by the University of North Texas (UNT) and National Taiwan Normal University (NTNU) from June 29 to July 1, 2020 and was supposed to take place at the UNT campus. Around early February, the preparation for the event was almost done. The flight tickets for the keynote speakers and the invited speakers were already booked; the conference venues, the hotels, the on-campus accommodations, and the catering plan were all set. At the early stage of COVID-19, in February 2020, the most affected areas were China and Europe. The PPTELL 2020 committee paid close attention to the development of COVID-19 but believed that it would be alright in the U.S. However, from March 2020, the spread of COVID-19 reached the U.S., which then imposed a shutdown and a curfew. Then, the PPTELL 2020 committee received the announcement made by UNT on March 22, indicating that large gatherings were prohibited. The worst was that the closure might continue for the whole spring semester and summer in 2020.

The bolt from the blue urged the committee to decide whether to cancel PPTELL 2020. If not, how could we host PPTELL 2020 and provide the attendees with a safe opportunity for successful interpersonal connection and experience sharing under the COVID-19 pandemic?

1.2. Moving PPTELL 2020 online as a virtual international conference

A thorough discussion of the abovementioned issues was conducted to help the committee make a better and appropriate decision. The discussion covered the budget, the potential and inevitable loss, the approaches to dealing with the cancellation of the already-booked items such as the keynote speakers' flight tickets, the proceedings publishing plan which was under the contract between the editorial team and Springer, the approaches to reaching the general purposes of an international conference and the expectations of the participants and attendees of PPTELL 2020, etc. After a careful analysis of the pros and cons, the committee decided to continue hosting PPTELL 2020 on the same dates by transforming it into an online conference. An urgent question following the decision needed to be answered was "Which platform should be chosen for the online conference?" It must be an easy-to-use, stable and secure one. Additionally, it should be affordable and should not cost more than that originally planned for hosting the conference. Based on the following reasons and the other conference organizers' experience (e.g., Misa et al., 2020), Zoom was chosen as the platform of PPTELL 2020 although the other suitable online platforms are also available: (1) given Zoom has been widely used for synchronous meetings in different domains (Wiederhold, 2020), such as business and education, it was supposed that most of the participants were familiar with the platform; (2) several registered participants were based in China and they might have difficulty using some other online platforms, such as Google Meet; (3) Zoom allows the conference to accommodate a large group of participants; (4) Zoom provides participants with two-mode chat function, both private and public; and (5) Zoom allows for recording and automatically transcribing the meeting.

One basic but serious challenge accompanied this decision: How to prevent Zoombombing (Marotti, 2020) during the conference? To learn about the real situations of a Zoom conference, one of the committee members attended an international conference on Zoom after it was chosen to run PPTELL 2020. She experienced an example of the unwanted Zoombombing, filled with endless bloody and pornography bombing, during the conference. Consequently, that meeting was forced to be terminated not long after its beginning. The unpleasant experience shocked the committee of PPTELL and reminded them that a series of actions had to be adopted to avoid similar attacks during PPTELL 2020.

1.3. The challenges

As mentioned in the previous section, the committee of PPTELL made a quick decision on responding to the influence caused by the COVID-19 pandemic, i.e., replacing the physical PPTELL 2020 with the virtual PPTELL 2020 on Zoom. The goal of PPTELL 2020 was the same as that of the original one: to reach a resounding success and make a valuable contribution to the academic community despite the Zoombombing threats. Additionally, more challenges emerged once Zoom was chosen as the platform of PPTELL 2020. Some of the challenges were also discussed with the organizers or hosts of other conferences, such as the 2020 edition of the Passive and Active Measurement (PAM) conference (Misa et al., 2020). Table 1 lists the challenges faced by the committee in different stages, i.e., pre-, during, and post-conference.

Table 1. A list of challenges in different stages

Conference organizing stages	Possible challenges encountered
Pre-conference	• How to arrange a program that is appropriate for the attendees in different time zones? • How to include all the essential elements of an international conference in an online one, such as the opening and the closing ceremonies, and the Q&A sessions following a speech or a paper presentation? • How to organize the team and assign duties to cover all the tasks needed to be completed in an online conference?
During conference	• How to guarantee the program goes smoothly and safely? • How to reach the speakers or the presenters if they do not show up? • How to keep the meeting going when the key persons, such as the speakers and presenters are unexpectedly disconnected during the meeting? • How to build an inter-personal connection among the attendees as it is expected in a physical international conference?
Post-conference	• How to build an inter-personal connection among the attendees as it is expected in a physical international conference? • How to obtain the participants' perception and evaluation on PPTELL 2020?

The challenges listed above concerned the following four main categories: the security issue in a Zoom meeting, the arrangement of the conference program, troubleshooting during the conference, and sustaining relationships between the participants, and collecting the participants' perceptions and evaluation on the conference. What's more challenging was that none of the main members of the organizing team had the experience in organizing an online international conference. Due to the lack of the practical experience of organizing an online conference, the team members learned by doing and constantly reflecting on what they did. The following sections briefly describe the targets and the corresponding actions to achieve them, as well as the feedback and comments given by the participants and attendees. Finally, a brief conclusion will provide the readers with suggestions for organizing an online conference.

2. The pre-conference stage

This stage included the steps to transform PPTELL 2020 from a physical conference to an online one. The main tasks in the pre-conference stage included (1) arranging an appropriate program to accommodate all the activities, (2) studying the functions of the Zoom platform, and (3) organizing and training a Zoom-monitor team.

Considering more than 100 attendees of PPTELL 2020 came from 12 countries/areas across ten time zones, arranging a program which included all the sessions allowing almost all the attendees to attend the program at a suitable time became the first challenge faced by the program committee. The first step was to confirm the number of attendees from each country/area and found that most of the attendees were located in the U.S. and Taiwan; some were in Canada and Chile, others were from Australia and New Zealand, and still others were from Japan, India, South Korea, Mainland China, Singapore, and Hong Kong. Next, the committee created a table of multiple time zones in which most attendees were located. Figure 1 shows the time table and the chosen time slots.

As shown in Figure 1, only 4.5 hours per day (marked in red) were appropriate for almost all the attendees, except one from India; thus, his session was arranged in the last hour within the 4.5 hours.

The second step was to squeeze all the events into the compact timetable, from at least 8 hours a day in a physical conference to 4.5 hours a day. The PPTELL 2020 events included the opening, the closing, four keynote speeches, two invited speeches, four workshops, 29 paper presentations, and several breaks. Considering the limited time available each day and the rich events, four Zoom rooms were needed and only 15 minutes allowed for each paper presentation. The conference structure of PPTELL 2020 was similar to that of the T6SympoZOOM conference (Salomon & Feldman, 2020) except that the former conference involved many more activities than did the latter. Additionally, both types of the program of PPTELL 2020, the finalized program at a Glance and the detailed daily program, can be found on the official website of PPTELL 2020.

India+5.5	Taiwan, China, Singapore+8	Japan, Korea+9	Australian Eastern +10	New Zealand+13	USA CA-7	Canada Central-6	USA TX, MI-5	USA Mi, PA-4	Chile-3
22.5	1	2	3	6	10	11	12	13	14
23.5	2	3	4	7	11	12	13	14	15
24.5	3	4	5	8	12	13	14	15	16
1.5	4	5	6	9	13	14	15	16	17
2.5	5	6	7	10	14	15	16	17	18
3.5	6	7	8	11	15	16	17	18	19
4.5	7	8	9	12	16	17	18	19	20
5.5	8	9	10	13	17	18	19	20	21
6.5	9	10	11	14	18	19	20	21	22
7.5	10	11	12	15	19	20	21	22	23
8.5	11	12	13	16	20	21	22	23	24
9.5	12	13	14	17	21	22	23	24	1
10.5	13	14	15	18	22	23	24	1	2
11.5	14	15	16	19	23	24	1	2	3
12.5	15	16	17	20	24	1	2	3	4
13.5	16	17	18	21	1	2	3	4	5
14.5	17	18	19	22	2	3	4	5	6
15.5	18	19	20	23	3	4	5	6	7
16.5	19	20	21	24	4	5	6	7	8
17.5	20	21	22	1	5	6	7	8	9
18.5	21	22	23	2	6	7	8	9	10
19.5	22	23	24	3	7	8	9	10	11
20.5	23	24	1	4	8	9	10	11	12
21.5	24	1	2	5	9	10	11	12	13

*Note: **red**,* synchronous time; *green*, testing time 1; ***blue***, testing time 2; **24**, Time NG (unsuitable for arranging activities)

Figure 1. Finding the best meeting time across different time zones through a table

As mentioned in 1.3 The challenges, the committee was most concerned about the security issue during the conference that has been challenging many Zoom users (Secara, 2020). By carefully studying the functions and referring to the others' experience (e.g., Berkeley Information Security Office, 2020; Vigliarolo, 2020), the team identified some Zoom functions to ensure security in the Zoom rooms for online meetings as listed below.

- Waiting room (The Waiting Room feature allows the host to control when a participant can enter the meeting): Allow join with the host's confirmation
- Password required (For security's sake)
- Audio (To avoid influence from unexpected noise or background sound): Off when entering; Video (For security's sake): Required for identification
- Sound notification when someone joins or leaves (To avoid influence from unexpected noise or background sound): Host and co-host only
- Screen sharing and start sharing (To ensure all the presenters/chairs are allowed to share their screens with the participants and attendees): All participants
- Allow participants to rename themselves (For security's sake): Disable
- Co-host is needed (The Co-host feature allows the host to share hosting privileges with another user(s))

To handle the possible technical issues, a team in addition to the committee of PPTELL 2020 was needed. The team's missions were to help monitor the Zoom rooms during the conference to make sure the meetings go smoothly. Consequently, a volunteer team was formed. The criteria for recruiting the volunteers were that they must be capable of attending several training sessions. We did not, however, limit their nationalities or locations, since having volunteers from only certain parts of the world would not form any challenges as long as the volunteers could find mutual times for discussion. The team members were from the U.S., Taiwan, New Zealand, Singapore, and Mainland China. Four of them were set as the hosts of the Zoom rooms in advance while the others served as the co-hosts to co-monitor the meetings. Soon after the volunteer team was formed, a series of trainings on Zoom usage and monitoring meetings were arranged. The following tasks were necessary to hold Zoom meetings smoothly and all the volunteers were asked to practice repeatedly until they were familiar: check the attendees' identity through their full names and decide whether to permit their entry into the "room" instantly, monitor the progress of the meetings, and guarantee they went smoothly under a secured space.

Moreover, in addition to training the volunteer team, a tutorial on using Zoom, especially the screen sharing function, was also arranged for all the attendees. The tutorial information was sent to the attendees two weeks before the conference via a Google Form titled "The Survey of using Zoom in PPTELL2020" before the tutorial. Considering some attendees were unavailable, some videos for self-learning were also created and shared with the attendees.

During the volunteer training, a new issue raised: how to deal with the no-shows or an unstable Internet connection? "No-show" has been a serious issue existing in many international conferences (Gobert, 2020) while the quality of internet connection greatly influences the progress of an online meeting (Bonifati et al., 2020). To deal with abovementioned issues, three tasks were completed at the preparation stage. Firstly, all the attendees were asked to join a private group on a real-time social media for receiving all the information or reminders sent by the committee and most importantly for an easier contact during the conference. Consequently, rather than setting up an additional Slack channel as Misa et al. (2020) did, the committee chose WhatsApp in this conference since it was used by most of the attendees of PPTELL 2020. Secondly, the slides and videos of the speeches and the presentations were collected and shared with the Zoom-monitor teams. The team also tested all the collected materials beforehand. Lastly, two reminders were sent to all the attendees in the last week of preparation stage: the general information and the information about the Zoom rooms. The reminder with the general information was sent out twice, one week and five days before the event, respectively. This reminder emphasized: (1) when and how the reminders will be sent to the attendees, (2) a must that everyone joins the WhatsApp group with the REAL names, (3) a must that everyone joins the Zoom meetings with the REAL and FULL names, and (4) the conference program.

On the other hand, the reminder with the detailed Zoom login information was sent out to all the attendees two days before the conference. This reminder provided the passwords and the links to the Zoom rooms. It also emphasized that the login information must not be shared with others. Besides, it contained information about the opening and the closing ceremonies.

After a three-month preparation, the committee welcomed all the attendees online and started the 3-day adventure, which will be described below, the during-conference stage.

3. The during-conference stage

As listed in Table 1, a progress being smooth and secured are the two major elements at the during-conference stage. Although many organizers of virtual conferences shared their experience in 2020 (e.g., Rose, Mott, Alvarez, & Lin, 2020; Veldhuizen, Slingerland, Barredo, & Giller, 2020), some unexpected situations would possibly happen and provide the committee with new lessons to learn from. Therefore, the committee and the Zoom-monitor team wrote down what happened within the three days, especially the unexpected events.

In totally, there were 43 sessions in PPTELL 2020, including the opening ceremony, keynote speeches, invited speeches, workshops, panels, paper presentation, and the closing ceremony. The most-attended event was the opening ceremony; more than 80 people attended it, while the less-attended one was one paper presentation session with fewer than 20 attendees. Except for the opening ceremony, the plenary sessions which included the four keynote speeches attracted most of the attendees and participants. Regarding the security issue, the three-day events went without any problems. No any cases of zoombombing were found during the three-day events. But some common issues encountered by many other online conferences were also found in PPTELL 2020, such as the no-show issue and the poor quality of an Internet connection. Table 2 lists those issues, the frequencies and the percentages, and the descriptions of the matters.

It can be found that with the support of the Zoom-monitor team, the security aspect was not an issue during PPTELL 2020. However, although some videos were recorded beforehand, playing those videos, except those lasting within 2 minutes, did not go smoothly such as the welcome speeches given by the two Presidents of UNT and NTNU. What we experienced is different from that described in reports of Salomon and Feldman (2020) and Misa et al. (2020). According to the experiences obtained from this conference, therefore, it is suggested that the pre-recorded videos, except the very short clips, should be shared with the attendees and the participants in advance, rather than playing the recorded presentations during the conference. The limited session time should be used for interactive discussion, rather than for video watching, especially when the Internet quality at the presenter's side does not allow playing the video. Additionally, the issue of disconnection of the talks given by the invited speaker and the workshop organizer was resolved thanks to the session chairs' quick-witted action. Although the Zoom-monitor team's help did not work, the session chairs led the attendees and the participants through an interactive discussion over their research experience in the topics of those sessions. Furthermore, neither of the issues, the no-show and the platform usage, influenced the event much. With the support of the Zoom-monitor team, the three sessions mentioned in Table 2 turned out very smooth.

Table 2. The unexpected situations during the 3-day events

Unexpected matters	Frequency (percentage)	Event descriptions
No-show	2 (4.65)	1. One session chair who was also a presenter did not show up. 2. One presenter did not show up.
Platform usage (Zoom usage)	1 (2.32)	1. One presenter did not use the function of screen sharing appropriately at the beginning of the presentation.
The quality of internet connection	4 (9.30)	1. The pre-recorded videos of one panelist stopped during the session 2. The playing of a keynote speaker's pre-recorded videos was unstable. 3. One invited speaker disconnected at 15 minutes after the session started. 4. One workshop organizer disconnected 10 minutes before the session ended.

In sum, the overall activities of PPTELL 2020 went smoothly. According to the immersive observation, although some unexpected events occurred, the Zoom-monitor team played their role well. Additionally, the guidance given by the experienced session chairs helped resolve the issue of no-show and a poor Internet connection at the presenters' side.

4. The post-conference stage: Feedback from the participants and attendees

It was the very first experience for the committee to organize an online international conference as described in *section 1.3*. To make improvements based on the attendees' feedback, a post-conference survey was sent to all of the attendees via WhatsApp and emails. A total of twenty responses were collected. The results obtained from the questionnaire mainly fall into five categories: the participants' backgrounds, their opinions on the activities conducted before, during and after PPTELL 2020, and their overall satisfaction. Below is a brief description of the survey results.

4.1. Respondents' backgrounds

To learn more about the participants, the researchers included questions about the participants' locations while participating in PPTELL 2020, their roles in PPTELL 2020, research fields and their experience of using Zoom. The results show that during PPTELL 2020, 75% of the participants were in North America, while 20% of them were in Asia and 5% were in Oceania. When the participants filled in their roles in PPTELL 2020, they were allowed to select as many options, e.g., Chair, Presenter, Speaker and Attendee as they thought applicable from the list. Among the twenty responses, about 60% were provided by the presenters; the speakers and attendees each contributed about 16% to the responses, and one response was received from a session chair of PPTELL 2020.

Most of the respondents were researching either digital learning or language learning or both. Specifically, almost half of the participants who have submitted their responses were working on technology-enhanced language learning (TELL). About 20% of the respondents were researching e-Learning, followed by linguistics (17%), critical thinking (7%), information science (4%) and second language learning (3%). While 95% of the respondents have experienced using Zoom before attending PPTELL 2020, as few as 30% of respondents used Zoom for attending an online conference. Most of them used it for online meetings and online courses.

4.2. Comments on the actions taken by the committee before PPTELL 2020

Before PPTELL 2020, the organizing committee used multiple resources to disseminate information, to stay closely in contact with the speakers and participants and to provide solutions to any problems arising before the conference. These resources included a WhatsApp group, emails, Zoom tutorials, an electronic program booklet and a website.

Six questions concerning whether those ways of keeping the participants informed were helpful were answered with yes, no or not applicable. Over 80% of the respondents agreed that emails, the program booklet and the website were helpful, while only 60% and 35% of them considered a WhatsApp group and Zoom tutorials, respectively, as useful. Even though WhatsApp allows for instant messaging, some of the participants had some concern over revealing phone numbers. One respondent described it as “nosy,” probably due to the constant and sometimes personal updates running in the group. As high as 65% of the respondents chose not applicable when asked about the Zoom tutorials since few of them had participated in one, for they considered themselves familiar with the platform.

4.3. Comments on the events during PPTELL 2020

During PPTELL 2020, various events happened. To begin with, for security concerns, the participants had to wait in the virtual waiting room before they were granted access to the “room.” As many as 95% of the respondents were satisfied with such a decision and did not have to wait long, mostly for less than a minute. In terms of the length of the conference per day, i.e., 4.5 hours, some suggested it be slightly shortened. The organizing committee was originally planning to open extra rooms for the speakers and presenters to rehearse but later changed the decision to reduce complexity and avoid confusion. However, 25% of the respondents disagreed with their decision, stating that it would be better to have such an opportunity for rehearsing, for getting familiar with presenting works on Zoom and for resolving potential technical issues.

In general, PPTEL 2020 met the respondents’ expectations. They were satisfied with the quality of networking, though relatively limited in an online format. Particularly, the respondents enjoyed the exchange of knowledge and information and the “close” connection with each other. As an interdisciplinary conference, PPTELL 2020 brought together scholars from different yet related fields. It seems the participants liked the idea of gaining exposure to knowledge drawn from different disciplines which could then be linked to their original research fields. Results also show that the participants especially enjoyed the opportunity of “seeing” and “meeting” each other online, forming a genuine atmosphere. This feedback reflected the current situation affected by the pandemic when most people had to maintain physical distance, but at the same time seeking ways to stay in touch, intellectually and emotionally.

The respondents also provided valuable suggestions for handling technical issues. These suggestions can be divided into those that could be carried out by the committee and those by the presenters themselves. The organizing committee was praised for their timely assistance but was suggested to have a troubleshooting guide ready beforehand. On the presenters’ end, many of the recommendations point to the idea of having their talks pre-recorded as videos in case of heavy bandwidth load. This way of presenting works, however, drew different opinions. Thirty per cent of the respondents reported that such an approach should not be encouraged unless there was no any other option, while some took advantage of it, stating that the videos could be re-watched, which was useful when there was more than one session of interest at the same time. The respondents also appreciated the arrangement of the opening and the closing ceremony. They described these two events as short but necessary.

4.4. Comments on the follow-up connection after PPTELL 2020

The participants were also asked to indicate, based on the experience of attending PPTELL 2020, their willingness to tell the other scholars about PPTELL 2020 and to attend the next one online. Over 85% of the respondents provided positive responses. The reasons provided mainly center on great flexibility, mobility and connectivity, i.e., PPTELL 2020 on Zoom broke the barriers of time and space, which was essential especially during the COVID-19 pandemic. The participants were free from the risks imposed while taking flights or during face-to-face connections. Meanwhile, social networking was still made possible, though in a different form from what was often seen during an on-site conference, or in a physical space. The responses show that the quality of this new form of social networking was not compromised despite the difference.

4.5. Overall satisfaction with PPTELL 2020

The researchers designed twelve closed-ended questions on the overall satisfaction, including the time arrangement, each session, the selection of platform, (i.e., Zoom), etc. On a scale of 1 to 5, with 1 being very dissatisfied and 5 being very satisfied, the respondents rated each item. On average, all the twelve items received at least 4.00. The Opening ceremony received the highest score (4.90), while the arrangement of session breaks is rated the lowest (4.25). It seems the respondents were mostly quite satisfied with all the conference sessions, e.g., the keynote speeches, workshops, panels, paper presentations, etc., the rating of which reached 4.50 and above. On the other hand, the time arrangement for each day of PPTELL 2020 and the breaks was rated slightly lower. It was reported that PPTELL 2020 was quite intensive.

4.6. Interim summary

It can be found that the participants were satisfied with most of the events of PPTELL 2020, especially the Opening ceremony. However, the respondents also pointed out what the conference organizers should have paid attention to, along with some solutions to make the next online conference more successful and to reach more participants' expectation. Below are what should be considered when organizing the next online conference, in addition to what was described in the previous sections.

- There should be multiple platforms for contact and they should not involve the collection of any personal information, such as phone numbers.
- Although almost all of the respondents expressed that they were familiar with the Zoom platform and so they did not participate in the tutorials, some of the presenters and panelists were not familiar with Zoom according to what happened during the conference (see *3. The during-conference stage*). Therefore, a strong team of hosts/co-hosts should be organized to handle well when the presenters have problems handling the platform.
- A troubleshooting guide should be prepared beforehand for the event to go smoothly.
- A preparation room should be available for the presenters to get familiar with the platform and rehearse before their presentations.
- A longer break should be arranged to allow for more interaction between the participants or more rest after looking at the screens for too long.
- The length of the meeting time for each day should not be too long.

5. Lessons learned from organizing and hosting PPTELL 2020

After the 3-day adventure in organizing and hosting PPTELL 2020, some valuable lessons were learned from the experiences obtained from both stages, the during- and post-conference. Below briefly describes the lessons learned from this event.

- **Zoom hosts and session chairs:**

A well-trained team with dedicated members to serving as Zoom room hosts is necessary in addition to session chairs. The Zoom room hosts helped monitor the meeting, as well as deal with the unexpected situations. By doing so, the session chairs can focus on introducing the speakers/presenters and leading an interactive discussion. This requirement is also mentioned by Bonifati et al. (2020).

- **The length of a virtual meeting within a day:**

This seems to be one of the most concerned issue faced by most of the virtual conference organizers if they have participants from areas across multiple time zones (Wang, Vishwanath, Sitaraman, & Mareels, 2020). This conference took at most 4.5 hours for a one-day program, while the other conferences took longer (e.g., Bonifati et al., 2020; Rose et al., 2020) or shorter (e.g., Salomon & Feldman, 2020). It is found that the decision depends on both the numbers of the zoom rooms from which the participants came and the events accommodated in a conference.

- **The time allocated for each session, either a paper presentation or a keynote speech:**

A keynote speech was allocated one hour while each paper presentation was allocated 10 minutes in PPTELL 2020. Compared with the face-to-face and physical PPTELL conferences in 2018 and 2019, the length of the keynote speech in PPTELL 2020 remained the same due to the valuable experience shared by the speakers and the attendees' expectation. However, the time slots were shorter for paper presentations, compared with those at a regular conference due to the shorter conference duration. Many other conferences took the similar approach to accommodate more sessions in a virtual conference (e.g., Rose et al., 2020). Considering the two-way discussion among the attendees was necessary, additional time was allocated for the Q&A activity in each paper presentation session. The participants enjoyed the time arrangement and the opportunity for experience sharing and a discussion as they expressed in the post-conference survey (see Section 4.3).

- **The time for social networking:**

Due to the very tight schedule, there were only 10 minutes for the between-session break. It was found that although additional time was given for Q&A and the interactive discussion, there was still a need to arrange some time for social networking and refreshing. As described in Danaldson's (2020) report on the first virtual PLDI conference, many attendees would have appreciated some planned periods for gathering them for experience sharing. Similar feedback was also given by the attendees of PPTELL 2020 as described in Section 4. However, the attendees of PPTELL 2020 mentioned that the longer breaks would allow them to rest their eyes from staring at the screens for a long time.

- **The pre-recorded videos:**

Given that the quality of an internet connection is one of the major technical issues in an online environment (Rasheed, Kamsin, & Abdullah, 2020), some virtual conference organizers suggested that all the presentations should be pre-recorded and replayed during the conference (Bonifati et al., 2020; Misa et al., 2020). However, the suggestion mentioned above is not the case in what we learned from holding PPTELL 2020. Except the two pre-recorded welcome speeches given by the Presidents of UNT and NTNU, which went smoothly, all the videos recorded by the keynote speakers or the invited panelists failed during the conference, regardless of whether those videos were played by the speakers or the IT team of PPTELL 2020, which was located in a central place with a high capacity Internet connection. However, the session Zoom recordings did provide the attendees with an opportunity to review the event after the conference. Regarding the issue, the suggestion given here is to share the links to the pre-recorded videos with the attendees in advance. By doing so, the discussion can be led even if video playing is not smooth during the session.

- **The opportunity for professional development:**

PPTELL 2020 reached out more people than in the first two PPTELL conferences. In fact, a number of them could not attend the event if it was conducted physically at the UNT campus. This problem is in line with the experiences shared by many other virtual conferences, such as PLDI 2020 (Donaldson, 2020). Additionally, more graduate students participated in PPTELL 2020. It was also found that some presenters who are not native English speakers performed well without having to worry about their English oral skills. Presenting in English online seems to lower the presenters' anxiety. It echoes studies about foreign language in virtual environments (Wu, Lan, Huang, & Lin, 2019). Additionally, PPTELL 2020 functioned as the other virtual conferences in providing the attendees with a platform to enjoy the exchange of knowledge and information (Black, Crimmins, Dwyer, & Lister, 2020) and the "close" connection with each other as described in Section 4.3.

Figure 2. The goals and elements of online/virtual conferences

In sum, based on the lessons learned from organizing and hosting PPTELL 2020, the goals of holding online/virtual conferences should be knowledge and experience sharing, professional development, community building, and interpersonal connection. All the goals can be achieved only when a SAFE environment is created. As shown in Figure 2, in such an environment, the activities will go stably (stability); the information and expertise will be accessible (accessibility); the approaches to sharing knowledge and experience are flexible (flexibility); and all the events will be engaging (engagement).

A virtual conference seems to be a trend for academia now and in the future not only due to the COVID-19 pandemic but also for other considerations, such as decarbonization (Klöwer, Hopkins, Allen & Higham, 2020) and cost-saving (Bhargava, Farabi, Rathod, & Singh, 2020). Additionally, it amplifies social learning, professional development, and the role of technologies to enrich and expand the learning space (Spilker, Prinsen, & Kalz, 2020). The lessons learned from organizing PPTELL 2020 serve as the know-how for the next virtual international conference.

6. Conclusion

When "going online" becomes the most chosen or must-take option during the global pandemic, satisfaction, efficiency, and flexibility are as important as availability.

Organizing and hosting the PPTELL conference online was an unexpected plan for the committee, but the experience obtained from the process was beyond the challenges. It was appreciated that all the challenges turned into lessons and helped the committee get clearer about holding a satisfactory online conference. The feedback given by the participants and attendees through the post-conference survey also provides many insights for improving. It is of the authors' great pleasure to share the success and struggle of the PPTELL conference. What is shared in this paper can serve as a practical reference not only for the organizers/hosts of an online conference but also for online educators.

Acknowledgement

I want to thank the reviewers for their insightful comments and helpful suggestions. I also thank the Ministry of Science and Technology, Taiwan, ROC, under grant numbers MOST 106-2511-S-003-015-MY3 and MOST 109-2511-H-003-026- for financially supporting this research. I am also grateful that this research was partially supported by the Chinese Language and Technology Center of National Taiwan Normal University from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education in Taiwan.

References

Berkeley Information Security Office. (2020). *Settings for Securing Zoom.* Retrieved from https://security.berkeley.edu/resources/cybersecurity-and-covid-19/settings-securing-zoom

Bhargava, S., Farabi, B., Rathod, D., & Singh, A. K. (2020). The Fate of major dermatology conferences and meetings of 2020: Are e-conferences and digital learning the future? *Clinical and Experimental Dermatology, 45*, 759-761.

Black, A., Crimmins, G., Dwyer, R., & Lister, V. (2020). Engendering belonging: Thoughtful gatherings with/in online and virtual spaces. *Journal of Gender and Education, 32*(1), 115-129.

Bonifati, A., Guerrini, G., Lutz, C., Martens, W., Mazilu, L., Paton, Salles, M. A. V., Scholl, M. H., & Zhou, Y. (2020). *Holding a conference online and live due to Covid-19*. Retrieved from https://arxiv.org/pdf/2004.07668.pdf

Crawford, J., Butler-Henderson, K., Rudolph, J., Malkawi., B., Glowatz, M., Burton, R., Magni, P. A., & Lam, S. (2020). COVID-19: 20 countries' higher education intra-period digital pedagogy responses. *Journal of Applied Learning & Teaching, 3*(1), 9-28.

Cucinotta, D., & Vanelli, M. (2020). WHO declares COVID-19 a pandemic. *Acta Bio-medica: Atenei Parmensis, 91*(1), 157-160. doi:10.23750/abm.v91i1.9397

Cullinane, C., & Montacute, R. (2020, April). COVID-19 and social mobility impact brief #1: School shutdown. *The Sutton Trust, Research Brief.* Retrieved from https://www.suttontrust.com/wp-content/uploads/2020/04/COVID-19-Impact-Brief-School-Shutdown.pdf

Danaldson, A. F. (2020). *A Report on the first virtual PLDI conference*. Retrieved from https://arxiv.org/ftp/arxiv/papers/2007/2007.11686.pdf

Gobert, M. (2020). Conference organization as a path to professional development. In C. Coombe, N. J. Anderson, & L. Stephenson (Eds.). *Professionalizing your English Language Teaching* (pp. 281-290). Cham, Switzerland: Springer Nature Switzerland AG.

Hartshorne, R., Baumgartner, E., Kaplan-Rakowski, R., Mouza, C., & Ferdig, R. E., (2020). Special issue editorial: Preservice and inservice professional development during the COVID-19 pandemic. *Journal of Technology and Teacher Education, 28*(2), 137-147.

Klöwer, M., Hopkins, D., Allen, M., & Higham, J. (2020). An Analysis of ways to decarbonize conference travel after COVID-19. *Nature, 583*, 356-359.

Marotti, A. (2020, April 2). Zoom video meetings are being interrupted by hackers spewing hate speech and showing porn. It's called "Zoombombing." Here's how to prevent it. *Chicago Tribune*. Retrieved from https://www.chicagotribune.com/coronavirus/ct-coronavirus-zoombombing-20200401-wf2pvzqhbngitankuokvinvk2m-story.html

McKibbin, W., & Fernando, R. (2020). The Economic impact of COVID-19. In R. Baldwin, & B. W. di Mauro (Eds.), *Economics in the Time of COVID-19* (pp. 45-51). London, UK: Centre for Economic Policy Research. Retrieved from https://voxeu.org/system/files/epublication/COVID-19.pdf

Miró, J. D. V., Baquero-Arnal, P., Civera, J., Turró, C., & Juan, A. (2018). Multilingual videos for MOOCs and OER. *Educational Technology & Society, 21*(2), 1-12.

Misa, C., Guse, D., Hohlfeld, O., Durairajan, R., Sperotto, A., Dainotti, A., & Rejaie, R. (2020). Lessons learned organizing the PAM 2020 virtual conference. *ACM SIGCOMM Computer Communication Review, 50*(3), 46-54.

Rasheed, R. A., Kamsin, A., & Abdullah, N. A. (2020). Challenges in the online component of blended learning: A Systematic review. *Computers & Education, 144*. doi:10.1016/j.compedu.2019.103701

Rose, C., Mott, S., Alvarez, A., & Lin, M. (2020). Physically distant, educationally connected: Interactive conferencing in the era of COVID-19. *Medical Education, 54*(8), 758-768.

Salomon, D., & Feldman, M. F. (2020). The Future of conferences, today: Are virtual conferences a viable supplement to "live" conferences? *EMBO reports, 21*(7), 1-4.

Secara, I. (2020). Zoombombing- the end-to-end fallacy. *Network Security, 8*, 13-17.

Spilker, M., Prinsen, F., & Kalz, M. (2020). Valuing technology-enhanced academic conferences for continuing professional development. A Systematic literature review. *Professional Development in Education, 46*(3), 482-499.

United Nations Educational Scientific and Cultural Organization (UNESCO). (2020, March 4). *COVID-19 Educational disruption and response*. Retrieved from https://en.unesco.org/covid19/educationresponse

Veldhuizen, L. J., Slingerland, M., Barredo, L., & Giller, K. (2020). Carbon-free conferencing in the age of COVID-19. *Outlook on Agriculture, 49*(4), 321-329.

Verbeke, A. (2020). Will the COVID-19 pandemic really change the governance of global value chains? *British Journal of Management, 31*, 444-446.

Vigliarolo, B. (2020, April 3). *How to prevent Zoom bombing: 5 simple tips*. Retrieved from https://www.techrepublic.com/article/how-to-prevent-zoom-bombing-5-simple-tips/

Wang, D., Vishwanath, A., Sitaraman, R., & Mareels, I. (2020). *Organizing virtual conferences through mirrors: The ACM e-Energy 2020 experience*. Retrieved from https://arxiv.org/pdf/2008.08318v1.pdf

Wiederhold, B. K. (2020). Connecting through technology during the coronavirus disease 2019 pandemic: Avoiding "ZOOM fatigue." *Cyberpsychology, Behavior, and Social Networking, 23*(7), 437-438.

Wu, Y. J. A., Lan, Y. J., Huang, S. B. P., & Lin, Y. T. R. (2019). Enhancing medical students' communicative skills in a 3D virtual world. *Educational Technology & Society, 22*(4), 18-32.

Wu, J.-Y., Liao, C.-H., Cheng, T., & Nian, M.-W. (2021). Using Data Analytics to Investigate Attendees' Behaviors and Psychological States in a Virtual Academic Conference. *Educational Technology & Society*, *24* (1), 75–91.

Using Data Analytics to Investigate Attendees' Behaviors and Psychological States in a Virtual Academic Conference

Jiun-Yu Wu*, Chen-Hsuan Liao, Tzuying Cheng and Mei-Wen Nian
Institute of Education, National Chiao Tung University, Taiwan, R.O.C. // jiunyu.rms@gmail.com // abugu103@gmail.com // anke801211@gmail.com // iammo0715@nctu.edu.tw
*Corresponding author

(Submitted September 24, 2020; Revised October 28, 2020; Accepted December 13, 2020)

ABSTRACT: Amid the pandemic of coronavirus diseases, virtual conferences have become an alternative way to maintain the prosperity of the research community. This study investigated attendees' participatory behavior in a virtual academic conference (TWELF2020, Taiwan) and studied the interrelationship among their mastery experience, competence, and engagement to shed light on the development of virtual conferences. Data were collected based on 602 unique IDs via their unstructured trace data and 106 respondents to the post-conference questionnaire. Ten indices were derived from participants' unstructured log to describe the conference-based and session-based behaviors. Study results demonstrated that virtual conferences could facilitate the extended and deepened participation of the research community, nourish the participant-centered scholarship building, and create an engaging conference environment that reflects quality experiences regarding participants' mastery experience, competence, and engagement. The implications of the study can inform future virtual conference organization to provide more engaging and rewarding conference experiences for participants of all gender and academic ranks.

Keywords: Pandemic, Virtual conference, Mastery experience, Competence, Engagement

1. Introduction

Scholars' behaviors and learning activities change with the advancement and versatile uses of Information and Communication Technologies (ICT) (Luan et al., 2020; Sugimoto et al., 2017). Online conferences are one of the applications of technology advances, which can facilitate continuing professional development for geographically dispersed participants (Moore et al., 2016). However, the norm or the long-kept custom of professional meetings have largely remained its face-to-face tradition until there are inevitable reasons to change the routine practices. Notably, the pandemic of the Coronavirus Disease 2019 (COVID-19) has propelled the academic community to reconsider its style of the convention in order to avoid spreading or contracting the disease. Equally important is to sustain the development of the community's social and intellectual interaction (Wu, et al., 2020). Amid the uncertainty of the disease outbreak, orders to enforcing social distancing, stay-at-home, or shelter-in-place have become an international trend (e.g., Mervosh et al., 2020). Features such as live streaming, interactivity (e.g., raise a hand, text/voice chatting), and productivity (screen sharing, annotation, co-editing) have greatly improved the quality of human communication. Thus, turning face-to-face conferences into online or virtual conferences may be what the current technological tools (e.g., Microsoft Teams, ZOOM, WebEx, Google Meet, and Jitsi Meet) can do to contain the outbreak and to maintain the prosperity of the research community.

Nevertheless, there was a scarcity of studies investigating participants' behavior in professional face-to-face conferences (Jacobs & McFarlane, 2005). Even less researched was how participants adapt and react to the virtual forms of academic conferences (Moore et al., 2016). Notably, conferences in any form are an indispensable part in the professional development for people in the academia. In traditional face-to-face conferences, participants' demographic differences such as their gender and academic ranks significantly influence their perception about the conference as well as their behavior/ decision whether to stay or leave their academic career (e.g., Biggs et al., 2018; Látková et al., 2009). Therefore, it is worth investigating how these differences are associated with their virtual conference participation experiences and engagement. For the gender issue (e.g., Lee & Wu, 2013; Wu, 2014; Wu & Cheng, 2019), gender inequity was commonly reported in academic conferences, with female participants reporting more easily affected by perceived gender inequity in the conference (Biggs et al., 2018). On the other hand, conferences are great venues for student participants and faculty members/researchers to establish linkage and share expertise, experience, and innovation (Mata et al., 2010). However, the costly expenses of travel, accommodation, and registration may pose great challenges for early-career researchers to attend academic conferences (Henderson, 2015). Thus, investigating the association of participants' demographic characteristics with their perceptions and engagement in the virtual conference can contribute to the limited literature in virtual academic conferences and shed light on establishing a positive virtual environment to promote the professional academic development.

The study examines participants' behavior patterns in a one-day virtual conference in e-Learning via their anonymous log files (Wu et al., 2021). Their onymous reflection upon their perception, competence, and engagement regarding this virtual conference are also investigated via questionnaires. The virtual conference consists of several online events and meetings, including the opening and closing ceremonies, two keynote speeches, and five parallel meeting rooms consisting of 15 oral presentation sessions (three sessions in each meeting room). Based on the theory of self-efficacy (Bandura, 1993), we explore how participants' mastery experience (defined as satisfaction with previous successful experience) is associated with their competence in the virtual conference, which in turn is related to their engagement in the virtual conference. By studying the hypothesized association, we intend to inform future virtual conference organizers of the factors that are pivotal to attendee's participation experience and enhance the development of virtual scholarship. Thus, the purpose of this study was threefold. First, we intended to analyze participants' unstructured trace data from the online conference platform to understand the general state and the variability of their participation or presence in the virtual conference. Second, we examined participants' mastery experience, competence, and engagement in the online conference by analyzing their structured responses via the post-conference survey. Third, we aimed to provide suggestions and implications to enhance the scholarship in the technology-enhanced environment, with triangulating sources of evidence from the analytics of unstructured and structured participation data (Wu, 2020). Therefore, the research questions for this study are

- RQ1: What are the attendees' participatory behaviors in the virtual conference, as reflected by their anonymous trace data on the conference platform?
- RQ2: What is the state of participants' mastery experience, competence, and engagement in the virtual conference? Will participants' gender and their academic rank have a differential effect on their perceived mastery experience, competence, and engagement in the virtual academic conference?
- RQ3: What is the association among participants' mastery experience, competence, and engagement in the virtual conference?

2. Literature review

2.1. Human behavior in virtual academic or professional development activities

The capability of the technology enables the analysis and monitoring of participants' behavior as well as co-constructing knowledge in the virtual conference. For example, in a virtual conference, Moore et al. (2016) showed that the percentage of participants who contributed to the text chat (active) ranged from 46% to 92% across six webinar sessions with an average active rate of 59.5%, which was regarded as significant participants' contribution. Moreover, the discourse types in the Webinars can be categorized into interpersonal (20%), evaluative (12%), technical (11%), procedural (8.5), or content (52%) based on the analysis of the text chats (Moore et al., 2016). Instead of moving face-to-face conferences online, some conferences used social media as a means for participating online and examined the effect of backchanneling on participants' behavior in the professional conferences (Kimmons & Veletsianos, 2016). Specifically, the expansion of technology-enhanced mobile and participatory online environments allows participants to have real-time texting or chatting when a presentation or event is ongoing, which is called "backchannelling" (Kellogg et al., 2006). Recent research reported that the social media-supported backchanneling alongside the main conference could enhance academic learning through expanded participation in conference programs and provide opportunities for more researchers to join the professional community (Greenhow et al., 2019). Early findings also revealed, "backchannel technologies empower members of the audience to communicate among themselves, and to investigate all kinds of related information and make these public" (p. 328, Jacobs & McFarlane, 2005). The study results support that these technological tools may promote user interconnections and decentralize the conference to uphold a co-constructed value in social scholarship in academia (Greenhow & Gleason, 2014). Like Webinars or backchanneling, virtual conferences can promote attendees' mutual communication via text chatting or direct audio and video streaming. However, they maintain the essential formats of their face-to-face version by moving their opening/closing ceremony, keynote speeches, parallel, or unparalleled oral presentations on the virtual space. The computer-mediated virtual conferences may exhibit similarities and differences as compared to face-to-face conferences. Thus, we would like to investigate how attendees adapt and react to the new style of conference presentation and learning to inform the design and organization of future virtual conferences, especially in a time of pandemic.

2.2. Attendees' perceptions, competence, and engagement

Ground on the theory of self-efficacy (Bandura, 1993), this study investigated attendees' perceptions, competence, and engagement in the virtual academic conference. Self-efficacy is key to one's self-regulation of motivation and is associated with engagement in the task as well as task performance (Bandura, 1977). Specifically, self-efficacy is the competence belief of what people think they can do. People may develop their self-efficacy based on four sources: mastery experiences, vicarious experiences, verbal persuasion, and emotional and physiological states, among which mastery experience is posited as the most potent source (Bandura, 1997). Empirical studies provided evidence to demonstrate the interrelationship among mastery experiences, self-efficacy, and task engagement. For example, mastery experiences (operationalized as a sense of satisfaction with one's past teaching success) were positively related with teachers' self-efficacy beliefs (Lee et al., 2019); moreover, prior teaching success weighed more for novice teachers' self-efficacy due to their limited mastery experiences compared with the experienced teachers (Tschannen-Moran & Hoy, 2007). Additionally, in a sample of 595 primary and secondary school teachers, teachers' self-efficacy was positively associated with changes in their work engagement (Granziera & Perera, 2019). Moreover, surveying 252 undergraduate and graduate students about their sources of Internet self-efficacy, Chuang, Lin, and Tsai (2015) showed that prior successful experiences of Internet use played an essential role in participants' Internet self-efficacy. Researchers also revealed that general Internet self-efficacy predicted more informational Internet activities, especially among non-experts in technology (Jokisch et al., 2020).

Based on the relevant studies about mastery experiences, self-efficacy, and engagement, we postulated that attendees' mastery experiences would be correlated with their perceived competence in virtual conferences, which would, in turn, predict their engagement in the virtual academic conference.

2.3. Gender and academic rank differences in the perception of academic conference participation

Professional conferences are great venues for scholars to share their research, obtain the latest information/developmental trend in their field, and communicate with fellow scholars within the same or across different research fields. However, gender issues exist in the presentation and participation in academic conferences. Jones et al. (2014) reported that women consistently presented less time than their counterparts regardless of their academic rank in a conference that had a 1:1 gender ratio. Women also asked fewer questions than men (1:1.8) in a scientific conference that promotes a clear code of conduct in prohibiting any form of discrimination (Hinsley et al., 2017). The perception of sexism or gender inequality in conference participation may have a detrimental effect on women's career development. Mainly, Biggs, Hawley, and Biernat (2018) showed that women who felt sexism and silenced at the conference would increase their intention to leave academic careers while men who perceived sexism would increase their intention to leave that specific conference but not the academia. In this study, we would investigate the gender differences in attendees' mastery experiences, competence, and engagement in the virtual conference.

Moreover, membership and participation in academic conferences are an essential means of professional development for both student and professional participants (Mata et al., 2010). In particular, students reported that interacting with professors or researchers was goal attainment in their academic development (Cheng et al., 2019; Látková et al., 2009). Moreover, participating in academic conferences can generate a research culture among students (Hall, 2015) and have positive impacts on studies or career, presentation skills, personal confidence, as well as research skills and perspectives (Little, 2020). However, it is not known how differences in academic rank will impact participation in virtual academic conferences. We would examine individual differences in academic ranks regarding mastery experiences, competence, and engagement in the virtual conference.

3. Method

3.1. Data source

The current study included two sources of attendees' data in a virtual conference of e-Learning in Taiwan (Taiwan e-Learning Forum of year 2000, TWELF2020), namely the unstructured behavioral data collected from the Zoom conferencing platform and the structured assessment of attendees' virtual conference experience via the post-conference questionnaire. For the past 14 years, the conference was held annually in late March as face-to-face conferences. However, it was transformed into a virtual conference due to the pandemic of COVID-19 in 2020.

The post-conference survey link was sent to 150 people who registered the conference, of which 106 provided their full response to the questionnaire (response rate = 70.67%) because all the question items were set as required. 72.2% of respondents have experience in attending the previous face-to-face meetings. 44.3% of the respondents were female, and 55.7% male. Among the respondents, their registration status can be categorized into session chair (9.4%), presenter (72.6%), and participants (17.9%). Their academic rank can be classified into student participants (e.g., master and doctoral students: 47.2%) or professional participants (e.g., researchers, scientists, faculty members: 52.8%).

3.2. Measures

3.2.1. Unstructured data of log traces

The unstructured data was collected from the log or traces of participants logging in the Zoom conference platform. Room A was the main virtual conference venue that hosted three activities in the morning: the opening ceremony and two keynote speeches. In the afternoon, three consecutive parallel sessions (i.e., Oral Paper session 1~3) were hosted in five parallel virtual meeting rooms (i.e., Room A, B, C, D, and E). We can only get the usage data from the first three rooms (i.e., Room A, B, and C) across three sessions of presentation due to the limited reporting features in the free subscription licenses. The closing ceremony was then held in Room A as the last event of the conference. We analyzed unstructured data on the conference-based and session-based units. The conference-based behaviors included three indices: (1) the total number of log-ins, (2) the total number of participants (calculated as the sum of unique IDs), (3) the instant maximum number of participants across sessions. The session-based behaviors consisted of nine indices: (1) average number of participants in each session, (2) instant maximum number of participants in each session, (3) session duration, (4) average participation duration, (5) average percentages of participation, (6) the number of dedicated participants (defined as the number of participants who completed a session for at least 70% of the time), (7) audience rating: an index score considering proportion of viewing time in the time period of a specific session over the total number of conference participants in the given time as suggested by Meyer and Hyndman (2006), (8) popularity rating: percentage of dedicated participants in each session (i.e., # of dedicated participants over the number of participants in each session), (9) retention rate (i.e., the ratio of common IDs stays at same room for the next session), and (10) the average number of switches in parallel sessions (i.e., the times that a participant leaves a room and joins another room in parallel sessions).

3.2.1. Structured data of psychological measurement

The structured data was collected using the researcher-developed questionnaire. The questionnaire was comprised of two parts. Part I collected participants' demographic information, such as their gender, age, and academic rank. Part II asked their mastery experience, competence, engagement, and general perception of the conference.

Mastery experience was adapted from Tschannen-Moran and Hoy (2007) and operationalized as participants' satisfaction in their experiences of participating in keynote speeches, oral presentations, and overall conference (3 items). Responses were rated on a 5-point Likert scale, with 1 *extremely disagree* to 5 *extremely agree*. Sample item is "I am satisfied with my overall experience in this virtual conference." The standardized factor loadings ranged from .62 to .89 of this just-identified Confirmatory Factor Analysis (CFA) model (Wu et al., 2018). Internal consistency was .81, with AVE = .82 & CR = .62.

Competence in the virtual conference was developed by adapting Harter's perceived competence scale (1982). The developed scale had three dimensions: social interaction competence (3 items), academic competence (3 items), and ICT use competence (4 items). Social interaction competence reflected participants' perceived competence to interact with new or familiar peers in academia. Academic competence assessed their perceived competence in presenting, receiving, and sharing academic findings. ICT competence demonstrated their perceived competence in using technology to prepare, join, or switch between different presentations or media. Sample items included "I am certain I can make new friends in the virtual conference (social interaction)," "I believe I can obtain the latest research development or trends in the virtual conference (academic)," "I am confident that I can switch between rooms and attend more parallel sessions in the virtual conference (ICT)." Responses were rated on a 5-point Likert scale, with 1 *extremely disagree* and 5 *extremely agree*.

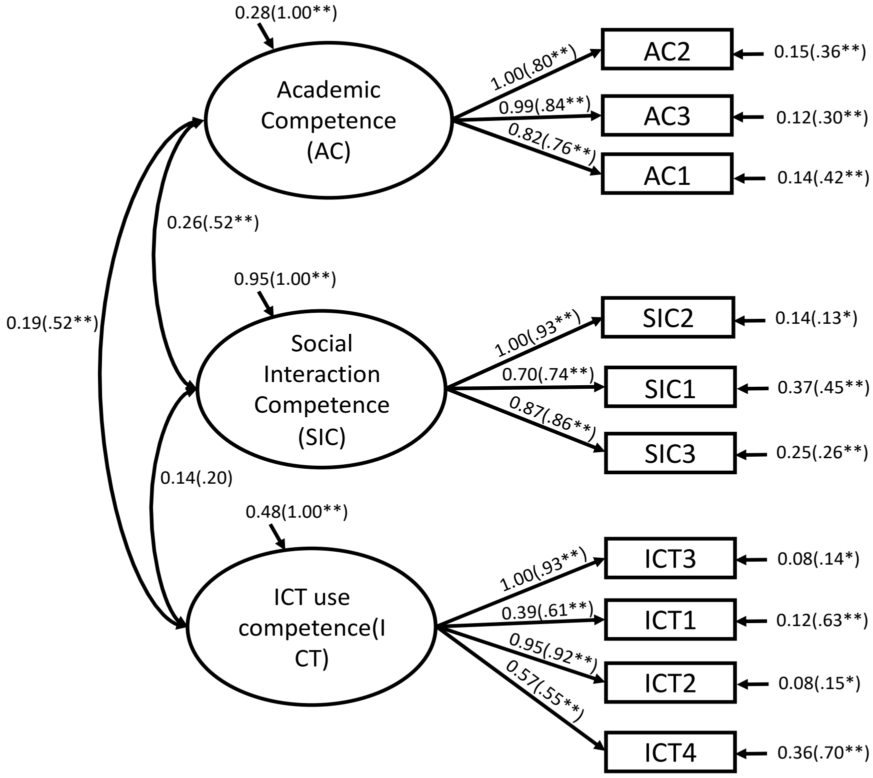


Figure 1. Three-factor Confirmatory Factor Analysis (CFA) for Competence in Virtual Conference Questionnaire (*Note.* Model-fit information: $\chi^2 = 39.75$, $df = 32$, $p = .16$, $CFI = .98$, $TLI = .98$, $RMSEA = .05$, $SRMR = .06$. The values in parentheses were standardized coefficients. $^{*}p < .05$; $^{**}p < .01$)

As shown in Figure 1, a three-factor CFA was fitted to the competence belief data. The model indicated adequate fit to the data, $\chi^2 = 39.75$, $df = 32$, $p = .16$, $CFI = .98$, $TLI = .98$, $RMSEA = .05$, $SRMR = .06$. The standardized factor loadings ranged from .55 to .93. Internal consistency was .84 for academic competence, .88 for social interaction competence, and .83 for ICT use competence. The overall internal consistency was .84. Average variance extracted (AVE) ranged from .64 to .74 and composite reliabilities (CR) ranged from .84 to .89. The factor scores of the three constructs were saved as indicators for the latent factor of the virtual conference competence in order to test the structural relationship among mastery experience, competence, and engagement.

Participants' engagement was a behavioral measure quantified by the number of sessions they attended. Finally, we surveyed the general perception of the virtual conference. They responded to their preference in attending virtual or face-to-face conferences in the future and to their perceived engagement level of this virtual conference (i.e., more engaged, equally engaged, or less engaged).

3.3. Data analysis

We computed descriptive statistics to understand participants' behavior and perception in attending the virtual conference on the R platform (The R Core Team, 2020). Ten virtual conference behavior indicators from unstructured log traces were calculated from the author-built R package. As for the structured responses from questionnaires, we utilized lavaan package (Rosseel et al., 2019) with Full Information Maximum Likelihood estimation (FIML, Mehta & Neale, 2005) to perform CFA and SEM analyses (Wu et al., 2017). All univariate normality measures (kurtosis and skewness) were within ±6. We also performed the visual examination of Q-Q plots, which exhibited the relation between the expected value of normal distribution and the observed value (Hair et al., 2010; Kline, 2010). The measures and Q-Q plots suggested that the normality assumption held for the response variables. However, the multivariate normality measures (e.g., Mardia test) was statistically significant, which was commonly seen when the sample size was greater than 106 with more measured variables (Cain et al., 2017). Considering the possible data non-normality (Wu et al., 2014), we addressed the issue by applying the MLR procedure with Satorra-Bentler rescaled chi-square (χ^2) model fit test statistic and corrected fit indices (i.e., *CFI, TLI, RMSEA,* and *SRMR*) to evaluate the model goodness-of-fit (Hu & Bentler, 1999; Wu & Kwok, 2012). Besides, in order for the observed scores to be compared between groups on the same standing, we tested the measurement invariance of the competence scales in a series of models, including configural, metric, and scalar invariances across groups (Millsap, 2011). The instrument must demonstrate scalar invariance to reach valid conclusions regarding observed group differences (Wu & Cheng, 2019). Comparative models were regarded as statistically equivalent if $\Delta CFI \leq .02$(Cheung & Rensvold, 2002), $\Delta TLI \leq .05$ (Little, 1997), $\Delta RMSEA \leq .015$, and $\Delta SRMR \leq .01$(Wu & Hughes, 2015). Moreover, if the majority of criteria satisfy the suggested thresholds, measurement invariance assumptions are established (Wu & Hughes, 2015). Additionally, we tested the structural relationship among attendees' mastery experience, perceived competence, and engagement in a mediation analysis within the structural equation modeling framework (Chou & Lee, 2017; Wu, 2017).

4. Result

4.1. Descriptive statistics and correlations of unstructured and structured indicators

For the entire conference, there were a total of 1700 times of log-ins in record with 602 unique log-in IDs. Participants' log traces in the one-day virtual conference were visualized as a Gantt diagram in Figure 2.

Gantt diagram visualizes participants' traces regarding the conference rooms they visited over time. Each row indicates the participating pattern per attendee. For example, participant ID363, who switched conference rooms frequently, stayed in Room C from 14:20~14:30, in Room B from 14:30~14:45, then went back to Room C for a few minutes, and log out and return to Room A before he left the conference. For the three oral presentation sessions, more frequent changes in colors within the same session indicated more switches among presentation rooms.

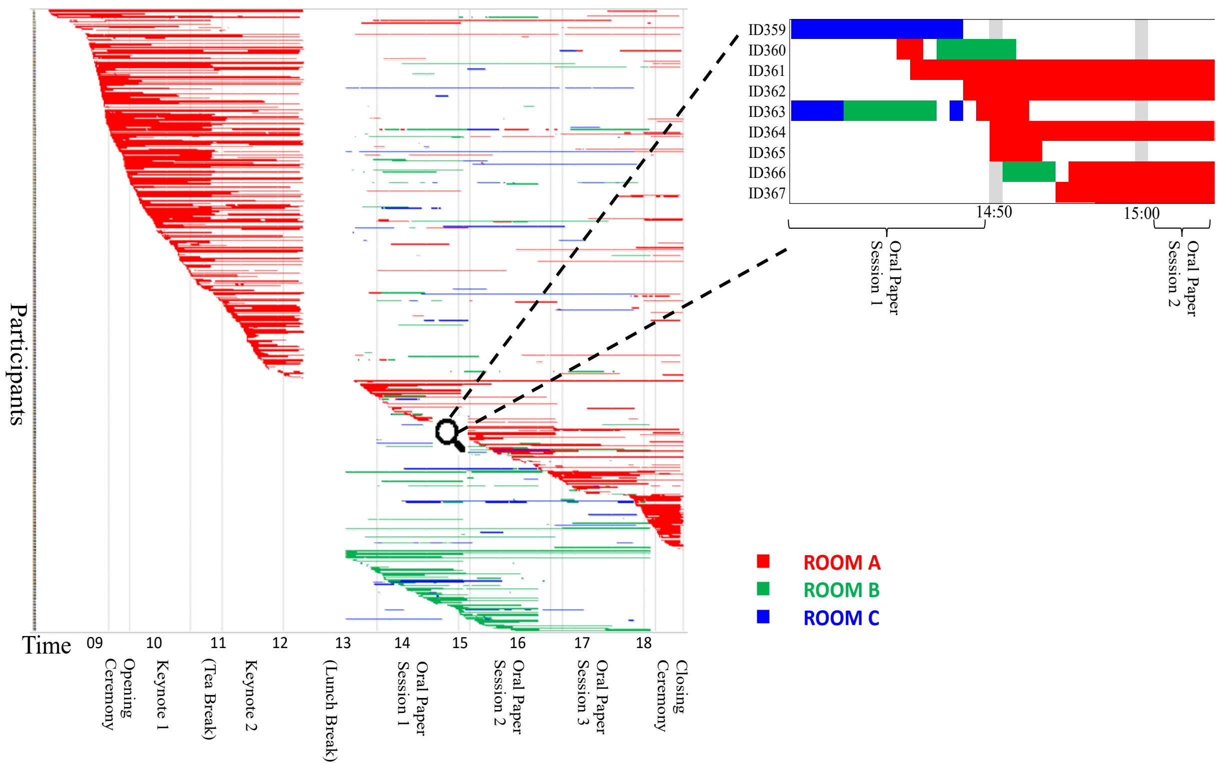


Figure 2. The Gantt diagram of participants' log traces in the one-day virtual conference. *Note.* Participants' traces were sorted in the order of starting time. Colors indicated different rooms that participants entered (Red: Room A, Green: Room B & Blue: Room C)

Figure 3 depicted the average (orange line) and instant (black line) number of participants from 8 am to 7 pm during the day of the conference. The average number of participants ranged from 75 to 130. The curvy black line reflected that the instant number of participants gradually increased before the start of each activity or session and dropped only a little bit during the session break, except for the lunch break. The instant maximum number of participants was 152 during the switch between the two keynote speeches.

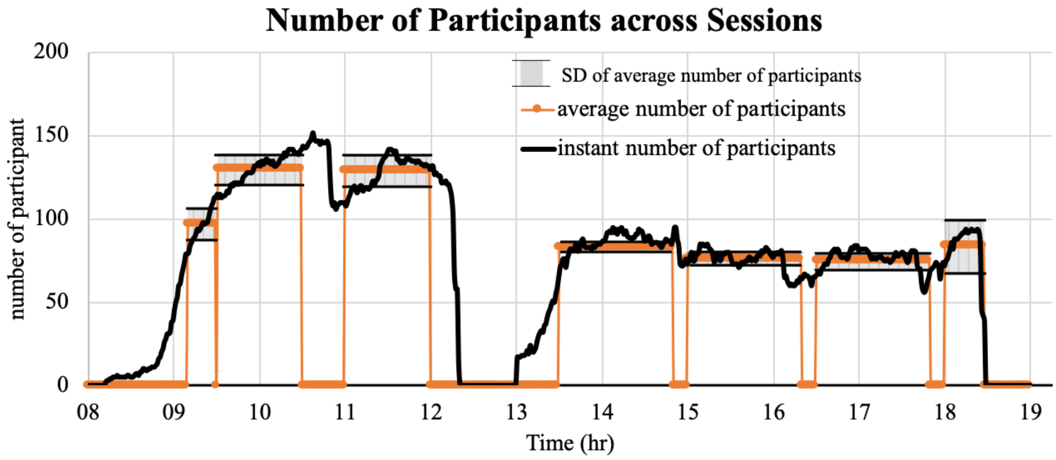


Figure 3. The average (solid orange line) and instant (black curve) number of participants from 8am to 7pm during one day conference. A gray area indicates the standard deviation of the number of participants in each session.

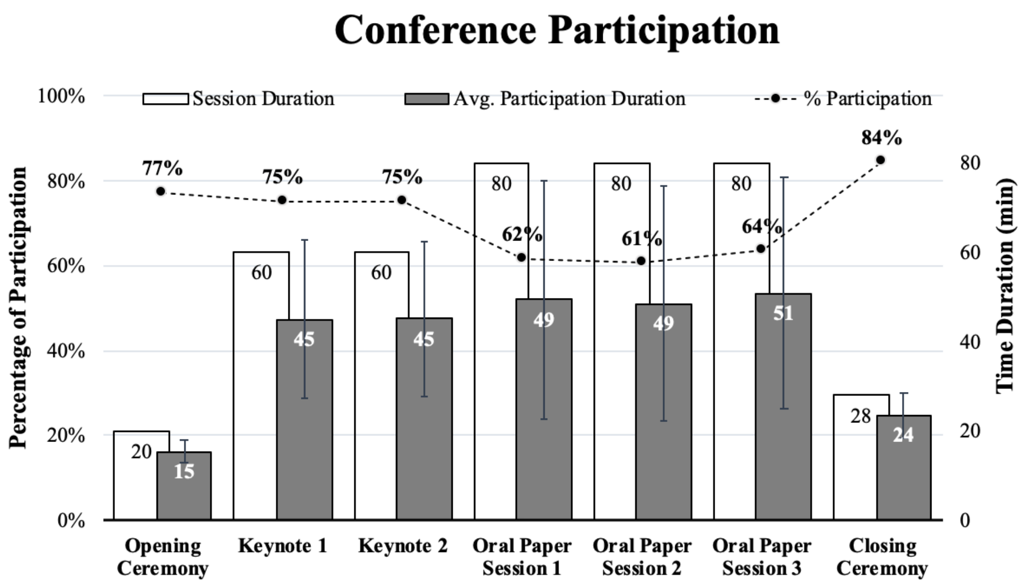


Figure 4. The session duration (white bar), average participation duration (gray bar with margin of error whisker), and average percentages of participation in each session

For the session-based behaviors, Figure 4 illustrated that attendees' average participation duration ranged from 15 to 51 mins within each activity or session. Proportional to session duration, average percentages of participation ranged from 61% to 84%. As shown in Figure 5, the number of participants who completed the session more than 70% of the time (dedicated users) ranged from 56 to 108. Thus, there was an audience rating of 9% to 18% of dedicated participants out of the total number of conference participants. Within each session, the average number of participants ranged from 75 to 130.

In terms of the popularity rating (percentage of dedicated participants in each session) as shown in Figure 6, the closing ceremony, the opening ceremony, and the two keynote speeches had the highest percentage of engaged participants, 87%, 86%, 83%, and 80% respectively. The average number of switches between parallel sessions ranged from 1 to 1.6 times. As for retention rate (i.e., the ratio of common IDs stays at the same room for the next session), keynote speech 1, opening ceremony, and presentation 3 had the highest retention rates, 95%, 92%, and 64%, respectively.

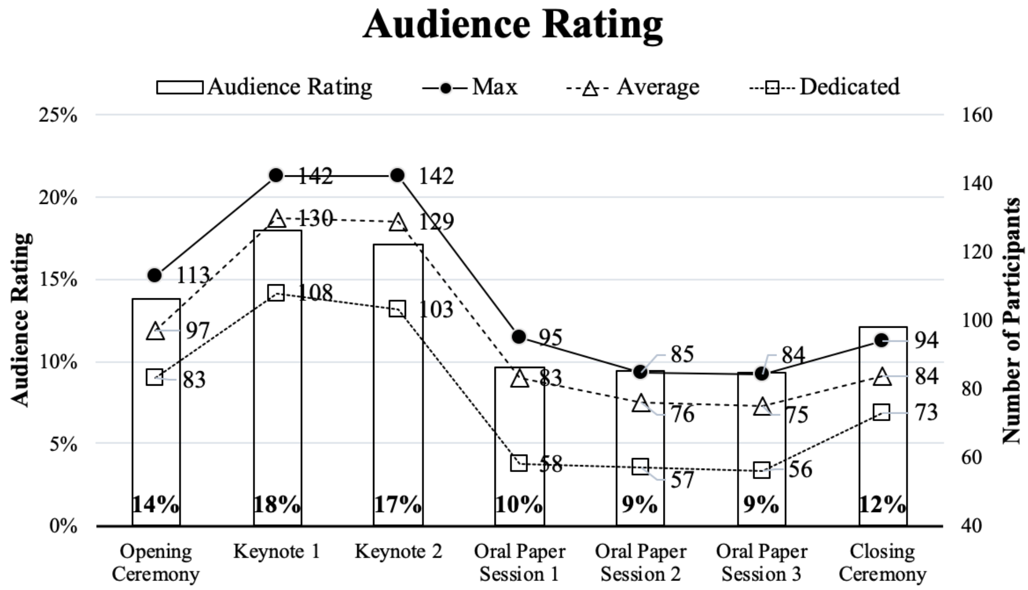


Figure 5. Audience rating (white bar with bold font) and the number of dedicated participants (# of participants who completed the session for more than 70% of the session duration) in each session, including the max and average number of participants

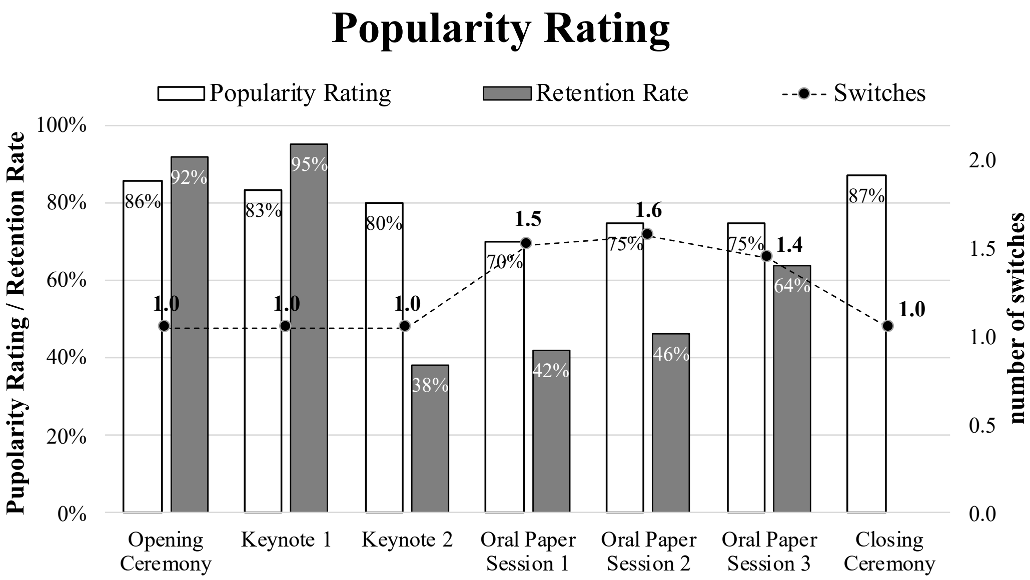


Figure 6. Popularity rating (white bar), retention rate (gray bar), and average number of switches (bold font) in each session

Descriptive statistics and correlations of structure survey responses were tabulated in Table 1. Attendees' average competence in attending the virtual conference was highest for ICT use competence, followed by academic competence, and social interaction competence (M = 4.58, 4.44, and 3.87, respectively). They exhibited high mastery experience in participating in the virtual conference (M = 4.68) and attended 4.10 meeting sessions on average. Mastery experience was positively related to all aspects of competences (r = .28~.60, p < .05). Academic competence was positively associated with social interaction competence and ICT use competence (r = .47 and .50, p < .05). Scores of ICT use competence, the overall competence, and mastery experience were positively correlated with the number of sessions participated (r = .28, .24, and .37, p < .05). 56.6% of the attendees reported their preference toward virtual conferences, while 43.4% reported favoring face-to-face conferences. Though more than half of the participants expressed that they were more engaged (11.3%) or equally engaged (45.3%) in the virtual conference compared with face-to-face conferences, 43.4% of respondents perceived virtual conferences to be less engaging.

Table 1. Descriptive statistics and correlations of structured responses

	1	2	3	4	5	6
1. Social interaction competence	--					
2. Academic competence	.47*	--				
3. ICT use competence	.19*	.50**	--			
4. Total competence	.78*	.80**	.71**	--		
5. Mastery Experience	.28*	.60**	.48**	.56**	--	
6. Engagement: # of sessions participated	.10*	.20	.28*	.24*	.37*	--
M	3.79	4.44	4.58	4.30	4.68	4.09
SD	0.89	0.54	0.55	0.49	0.42	2.80
Kurtosis	-0.25	-0.75	0.93	-0.38	0.10	-0.17
Skewness	-0.37	-0.47	-1.25	-0.36	-1.07	0.63

Note. $^{*}p < .05$; $^{**}p < .01$.

4.2. The measurement invariance tests

In order to compare the scale scores of competences in virtual conferences between groups, consecutive measurement invariance analyses were conducted. The configural assumption was first conducted to test the equality of the number of factors and the number of non-zero factor loadings across gender and academic rank. The results indicated the configural assumptions held for both gender and academic rank ($\chi^2 = 116.73$, $df = 64$ with $p < .01$, $CFI = .90$, $TLI = .85$ and $SRMR = .08$ for gender; $\chi^2 = 100.82$, $df = 64$ with $p < .01$, $CFI = .93$, $TLI = .90$ and $SRMR = .09$ for academic rank). Next, metric invariance was tested by further fixing the factor loadings equal across groups and was supported for both gender and academic status (Chi-square differential test $\Delta\chi^2 = 3.27$, $\Delta df = 7$ with $p = .86$, $\Delta CFI = .013$, $\Delta TLI = .023$ & $\Delta SRMR = .012$ for gender; $\Delta\chi^2 = 7.79$, $\Delta df = 7$ with $p = .35$, $\Delta CFI = -.002$, $\Delta TLI = .008$ $\Delta SRMR = .016$ for academic rank). Then, scalar invariance was tested by further fixing the item intercepts equal across groups and was also supported for both gender and academic status ($\Delta\chi^2 = 5.56$, $\Delta df = 7$ with $p = .59$, $\Delta CFI = .002$, $\Delta TLI = .015$ & $\Delta SRMR = .002$ for gender; $\Delta\chi^2 = 6.31$, $\Delta df = 7$ with $p = .50$, $\Delta CFI = .001$, $\Delta TLI = .010$ & $\Delta SRMR = .003$ for academic rank). The results of the MI analyses demonstrated that the measurement structure of competence beliefs was invariant across gender and academic ranks and can be directly compared with observed scores (Meredith, 1993).

4.3. Results of repeated measure ANOVA, the independent sample t-tests, and chi-squared test of independence

Due to high correlation coefficients among the competence measures, we conducted repeated-measures ANOVA to test if participants' three means of competence beliefs were equal. The Mauchly's test for sphericity was violated, W = .67, $p < .05$; thus, we adopted the Greenhouse-Geisser correction, $F = 48.96$, $p < .05$. Post-hoc test using Tukey contrast showed that participants' academic competence and ICT use competence were significantly higher than their social interaction competence ($\Delta M_{\text{Academic-Social}} = 0.65$ & $\Delta M_{\text{ICT use-Social}} = 0.79$, $p < .05$).

Table 2. The independent *t*-test on competence in virtual conference between gender

	Gender	*N*	*M*	*SD*	*t*	*p*
Social interaction competence	Female	35	3.75	0.74	-.32	.75
	Male	48	3.81	0.98		
Academic competence	Female	35	4.28	0.51	-2.41**	.02
	Male	48	4.56	0.54		
ICT use competence	Female	35	4.60	0.53	.22	.82
	Male	48	4.57	0.56		
Total competence	Female	35	4.25	0.48	-.84	.41
	Male	48	4.34	0.50		
Mastery Experience	Female	47	4.62	0.45	-1.24	.22
	Male	59	4.72	0.40		
# of sessions participated	Female	47	3.77	2.49	-1.08	.28
	Male	59	4.36	3.02		

Note. Tukey contrast post-hoc test was used. $^{*}p < .05$; $^{**}p < .01$.

The results of independent sample t-tests were shown in Table 2 and Table 3 for gender and academic rank. We observed a gender difference in academic competence, where women exhibited lower academic competence than men (M_{female} = 4.28, M_{male} = 4.56, t = -2.41, p = .02). Compared with student participants, professional participants had higher mastery experience ($M_{student}$ = 4.56, $M_{professional}$ = 4.78, t = -2.79, p = .01) and participated in more sessions ($M_{student}$ = 3.34, $M_{professional}$ = 4.77, t = -2.70, p < .01) in the virtual conference.

Table 3. The independent t-test on competence in virtual conference between academic rank

	Academic rank	N	M	SD	t		p
Social interaction competence	Student	41	3.88	0.82	-.92		.36
	Professional	42	3.70	0.95			
Academic competence	Student	41	4.33	0.60	-1.90		.06
	Professional	42	4.55	0.46			
ICT use competence	Student	41	4.69	0.49	1.74		.09
	Professional	42	4.48	0.59			
Total competence	Student	41	4.34	0.52	-.64		.52
	Professional	42	4.27	0.46			
Mastery Experience	Student	50	4.56	0.47	-2.79	**	.01
	Professional	56	4.78	0.34			
# of sessions participated	Student	50	3.34	2.59	-2.70	**	< .01
	Professional	56	4.77	2.83			

Note. Tukey contrast post-hoc test was used. $^{*}p$ < .05; $^{**}p$ < .01.

We tested the association of gender and professional rank difference on participants' preference toward virtual or face-to-face conferences and their perceived engagement with the chi-square test. The results were tabulated in Table 4 and Table 5. Gender did not exhibit an association with the preference (χ^2 = .30, df = 1, p = .58) but academic rank did (χ^2 = 9.13, df = 1, p < .01). Student participants preferred participating in virtual conferences while professional participants preferred participating in face-to-face conferences (adjusted standardized residual = -3.0, respectively). As for perceived engagement level in the virtual conference, we found disproportionately more counts of women reported being more engaged in the virtual conference, while disproportionately fewer counts of men were also found reporting more engaged in the virtual conference (χ^2 = 8.55, df = 2, p = .01, adjusted standardized residual = ±2.9). Similarly, we found more counts of professional participants reported being more engaged in the virtual conference than student participants (adjusted standardized residual = ±2.9); besides, more observed student participants than professional participants reported equally engaged (adjusted standardized residual = ±2.1) in the virtual conference compared with the face-to-face conference (Pearson χ^2 = 9.71, df = 2, p < .01).

Table 3. Chi-square test of conference preference between demographic variables

Conference preference		Gender		$\chi^2(1)$
		Female	Male	
Virtual	Count	28	32	.30
	Expected Count	26.6	33.4	
	Adj. Standardized Residual	0.6	-0.6	
Face-to-face	Count	19	27	
	Expected Count	20.4	25.6	
	Adj. Standardized Residual	-0.6	0.6	
		Academic rank		$\chi^2(1)$
		Professional	Student	
Virtual	Count	24	36	9.13**
	Expected Count	31.7	28.3	
	Adj. Standardized Residual	3.0	-3.0	
Face-to-face	Count	32	14	
	Expected Count	24.3	21.7	
	Adj. Standardized Residual	-3.0	3.0	

Note. Pearson χ^2 test statistics with degrees of freedom in paratheses was reported. $^{*}p$ < .05; $^{**}p$ < .01.

Table 4. Chi-square test of perceive engagement between demographic

Perceived engagement		Gender		$\chi^2(2)$
		Female	Male	
More engaged	Count	10	2	8.55**
	Expected Count	5.3	6.7	
	Adj. Standardized Residual	2.9	-2.9	
Equally engaged	Count	20	28	
	Expected Count	21.3	26.7	
	Adj. Standardized Residual	-0.5	0.5	
Less engaged	Count	17	29	
	Expected Count	20.4	25.6	
	Adj. Standardized Residual	-1.3	1.3	
		Academic rank		$\chi^2(2)$
		Professional	Student	
More engaged	Count	11	1	9.71**
	Expected Count	6.3	5.7	
	Standardized Residual	2.9	-2.9	
Equally engaged	Count	20	28	
	Expected Count	25.4	22.6	
	Standardized Residual	-2.1	2.1	
Less engaged	Count	25	21	
	Expected Count	24.3	21.7	
	Standardized Residual	0.3	-0.3	

Note. Pearson χ^2 test statistics with degrees of freedom in paratheses was reported. $^{*}p < .05$; $^{**}p < .01$.

4.4. Association among mastery experience, competence, and engagement

In order to test if participants' virtual conference experience is in line with the self-efficacy theory (Bandura, 1993), a structural model was fitted among mastery experience, competence, and engagement in the virtual conference, where competence mediated the association between mastery experience and engagement (i.e., the number of Zoom meetings attended) (Wu & Peng, 2017).

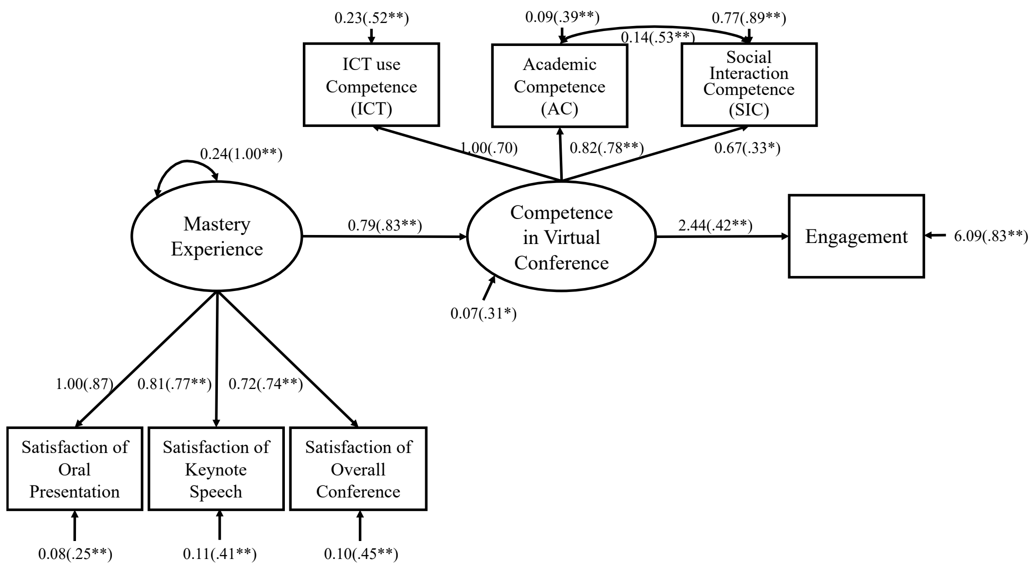


Figure 7. The structural model among mastery experience, competence, and engagement in the virtual conference. *Note.* Model-Fit Information: $\chi^2 = 16.24$, $df = 12$, $p = .18$, $CFI = .98$, $TLI = .97$, $RMSEA = .07$, $SRMR = .05$. $^{*}p < .05$; $^{**}p < .01$.

The hypothesized model had an adequate fit to the data, $\chi^2 = 16.24$, $df = 12$, $p = .18$, $CFI = .98$, $TLI = .97$, $RMSEA = .07$, $SRMR = .05$). The analysis results were illustrated in Figure 7. As expected, mastery experience positively predicted competence of virtual conference participation ($\beta_{\text{mastery}\rightarrow competence} = .83$, $p < .01$), which in turn was associated with more Zoom sessions attended ($\beta_{\text{competence}\rightarrow\text{engagement}} = .42$, $p < .01$). The standardized indirect association of mastery experience with engagement was .35 via competence of virtual conference participation ($\beta_{\text{mastery}\rightarrow\text{competence}\rightarrow\text{engagement}}$ = .35, $SE_{\text{Sobel test}}$ = .57, t = 3.35, p < .01). The variance explained R^2 was 69% for competence and 17% for engagement.

5. Discussion

This study adds to a burgeoning, yet a scarce body of literature that investigates virtual conference participation behavior (e.g., Moore et al., 2016). With the more frequent implementation of virtual conferences due to the convenience of use or to the avoidance of disease outbreak, the growing research may strengthen the development of a more comprehensive corpus of an empirical base for understanding virtual conference experience and behavior. The overarching goal of our research is to answer the three questions: (1) To what extent do attendees participate in the virtual conference as reflected by their trace data on the conference platform? (2) What are the overall and individual difference in participants' mastery experience, competence, and engagement? (3) What is the association among attendees' mastery experience, competence, and engagement in the virtual conference? In the age of pandemic, social distancing has become a norm. More professional conferences will be held in the virtual form. Besides virtual conferences' advantages in budgets and benefits to the environment (e.g., less travel and less pollution), this study used a data driven approach to explore attendees' participation pattern and their perceptions about the participation experiences. Findings were discussed in the terms of three following themes: extended and deepened participation, individual differences in virtual conference participation, and the association among virtual conference perception, competence, and engagement.

5.1. Extended and deepened participation experiences in the virtual conference

Analyzing the participation behavior via the trace data on the platform, we discovered extended participation for learning in the virtual conference. Specifically, due to the pandemic of COVID-19, this e-Learning conference was transformed into a virtual one and was open to people around the world with access to the ZOOM meeting links. As a result, there were 150 registered participants on the official record, but we obtained 602 unique IDs with a total of 1700 times of log-ins. It was apparent that the open-access of the virtual conference increased the possibilities of participation from those who were not physically present in the meeting, achieving an effect similar to the backchanneling alongside a face-to-face conference (Greenhow et al., 2019). Moreover, unlike backchanneling where participants twitted mainly to reference the meeting or to promote scholarship and networking (Greenhow et al., 2019), we found that attending the virtual conference can enhance attendees' conference participation experience by allowing them to have full access to the meeting regardless of the physical constraints. For example, our findings revealed that the audience rating for the conference sessions ranged from 9% to 18%, suggesting that there were 56 to 108 dedicated participants in each session on average. Besides, the popularity rating also showed that there were more than 80% dedicated participants in several programs, such as opening ceremony, closing ceremony, and keynote speeches. Our retention rate analyses further indicated that programs such as opening ceremony and Keynote speech have successfully retained more than 90% of attendees who participated in the current program to join the next program. Research was scare in studying the retention rate in conference sessions. However, in the television viewing market, Jardine and Romaniuk (2009) reported a retention rate around 50-60% for primetime television viewing in Australia, which constitutes the majority of the audience size of the next program. Particularly, the quality of the program was the significant determinant of the lead-in audience retention (retaining audience from the previous program) (Jardine et al., 2016). The opening ceremony lasted for 15 minutes, followed immediately by the two keynote speeches (45 min each). Thus, the high retention rates were mostly due to the two keynote speeches, which were usually the most important talks delivered in academic meetings and featured the underlying theme of the conference as well as the latest research trends and scientific findings.

Additionally, we found that our conference participants switched across three rooms 1.5 times on average in a parallel session. Switching across conference rooms suggested that attendees left the meeting without staying in the same conference room until the session ends. It may also indicate more flexibility and control for the participants to choose the talks or presentations they were interested in. For example, participants may be interested in the 1st presentation in room A and the 2nd presentation in room B; thus, they may well switch between the two rooms upon finishing the 1st presentation in room A. In light of this perspective, virtual conferences may help decentralize the conference (Greenhow & Gleason, 2014) for more participant-centered scholarship building.

5.2. The overall and individual differences in conference perception, competence, and engagement

In general, participants demonstrated high mastery experience as well as high academic and ICT use competences in attending virtual conferences, while their social interaction competence was significantly lower. Compared with face-to-face conferences, participants attending virtual conferences via video conferencing may experience low social presence (Kreijns et al., 2011), while social presence was a strong positive predictor for learning satisfaction and performance (Richardson et al., 2017). Thus, technological and pedagogical strategies to enhance participants' low social interaction competence warrants more research.

In terms of individual differences in the virtual conference participation experience, our result exhibited a gender gap in academic competence and an association between gender and conference preference. Previous research showed that women presented less time or asked fewer questions than men in academic conferences (Hinsley et al., 2017; Jones et al., 2014). Similarly, our study revealed that women had less competence than men to present, share, or receive research findings than men in the virtual academic conference. Nevertheless, more than expected numbers of women reported being more engaged in virtual conferences than in face-to-face conferences. Virtual conferences may pose a naturally forming shield for some women. Thus, they can focus on presenting their research or participating in the presentation without worrying about the direct disturbance or judgment from others due to the reduced perception of shared space (Taylor, 2011).

Moreover, we observed academic rank differences in mastery experience, engagement, and conference preference. As an exploratory attempt to understand academic rank and conference participation, our results revealed that professional participants perceived higher mastery experience and attended more meetings than student participants in the virtual conference. Besides, in terms of the level of engagement, more than expected numbers of professional participants perceived that virtual conferences were more engaging than face-to-face conferences and that virtual conferences were equally engaging as face-to-face conferences. Meanwhile, more than expected numbers of student participants preferred attending virtual conferences, while more than expected numbers of professional participants preferred attending face-to-face conferences. As a well-known fact, duties for professional participants included attending academic conferences to present or receive the latest development in the field as well as building connections and networking with academic peers around the world. Thus, attending academic conferences is part of the "academic citizenship" (Macfarlane, 2007). These responsibilities can justify the higher mastery experience, more engaging experience, and more virtual conference meetings attended for professional participants. Professional participants' preference toward attending face-to-face conferences provided additional evidence to support their adherence to fulfilling their employment duties.

Notably, traditional conferences tend to deepen the division of social networking of participants with different backgrounds (De Vries & Pieters, 2007). Thus, professional participants are more prone to bond with their existing connections in face-to-face conferences, which, however, may reduce the value of academic conferences (Spilker et al., 2020). Nevertheless, Davidson and Lyon (2018) found that attending academic conference positively impacted undergraduate students' career aspiration and enhanced their sense of belonging to the academic community.

The study findings have profound implications for the conference organizers. Concerning the social constraints in face-to-face conferences, virtual conferences may emerge as technological tool to provide opportunities for networking with proper arrangement by the conference organizers, such as identifying influential people in the community (Wu & Nian, 2021) and supplying connections among attendees using conference management systems (Spilker et al., 2020). In addition, more detailed pre-conference instructions/materials can be delivered to participants (especially student participants) to assist their presentation or recommend presentation sessions to enhance their mastery experiences and engagement. Despite the distinct associations of gender and academic ranks with participants' perceptions and engagement, virtual conferences can be the best of two worlds in order for continuous social and intellectual interaction in the academia amid the pandemic of contagious diseases.

5.3. Association among perception, competence, and engagement

In line with Bandura's self-efficacy theory (Bandura, 1977; Bandura, 1996), the results of the structural model confirmed the association among attendees' mastery experience, competence, and engagement in the virtual conference. Specifically, we verified participants' sense of satisfaction with their past success in the virtual conference positively predicted their perceived competence in attending the virtual conference, which in turn predicted more meeting sessions attended (engagement) in the virtual conference. According to the theory, prior experiences provide the most reliable source of self-efficacy; particularly, past success in the task can strengthen the competence belief to hold out against temporary frustration or failure (Bandura, 1997). For example, in Tschannen-Moran and Hoy (2007), teachers' mastery experience positively predicted their self-efficacy. The association was stronger for novice teachers with fewer prior task success as opposed to the experienced teachers. Given this new form of virtual conferences, we believe that most attendees possess limited mastery experience in virtual conferences; thus, the association will hold for the general population in academia. Moreover, consistent with the well-established association between self-efficacy/competence and engagement (e.g., Granziera & Perera, 2019; Skaalvik & Skaalvik, 2016; Wu, 2017), we revealed a positive relationship between attendees' perceived competence in participating in the virtual conference activities and the number of virtual conference sessions they attended. Findings about the association between mastery experience and competence as well as competence and engagement led to the inference that the more satisfied attendees felt about their virtual conference participation was also related to their engagement in virtual conferences or the number of virtual sessions they attended. Mastery experience, together with competence, had a considerable effect size on engagement in the virtual conference. As an implication for conference organizers, more scaffolds, such as pre-conference instructions and definite program agenda with virtual session links, can be supplied to enhance attendees' mastery experience and their competence in attending virtual conferences.

6. Limitation and conclusion

Considering the pandemic around the world, holding conferences in the virtual form appears to be a viable solution to maintain the interactivity and productivity of the research community. More than half of the participants preferred attending virtual conferences, while the rest preferred attending face-to-face conferences in the future. Nevertheless, the study results should be interpreted in light of limitations. First, the study was conducted on participants in the e-Learning domain, whose attendees are prone to the application of innovative technologies in learning. Thus, the research findings may not be generalized to conference attendees in other fields. Second, the study included both unstructured trace data and structured survey data from participants to illustrate participants' explicit behavior on the conference platform and the implicit ratings of their participation experiences. The two sources of data, however, cannot be linked by participants' identities. Future research can be done to link the two data sources for a more comprehensive understanding of participants' conference experience. For example, attendees' trace data (e.g., % of time being present in the sessions) can be used to represent their "true" engagement in the structural model.

Despite the limitations mentioned above, findings of the current study further revealed that virtual academic conferences could have the potential to become the mainstream in organizing future conferences. In this study, we provided the indices from participants' unstructured log to describe their conference-based and session-based behaviors. We also developed the measurement tool of competence in virtual conferences with adequate psychometric properties to identify and compare participants' academic, social interaction and ICT competence about virtual conferences. Based on the analytical results, we demonstrated that virtual conferences could facilitate the extended and deepened participation of the research community (Greenhow et al., 2019; Jardine & Romaniuk, 2009), nourish the participant-centered scholarship building (Greenhow & Gleason, 2014), and create an engaging conference environment that reflects quality experiences regarding participants' mastery experience, competence, and engagement (Granziera & Perera, 2019; Tschannen-Moran & Hoy, 2007). Future research can be designed to test technological and pedagogical strategies that can provide participants a more engaging and rewarding conference experience, especially on refining their social interaction competence.

Acknowledgement

The authors are grateful to the co-chair and committee members of Taiwan e-Learning Forum 2020, including Professors Sunny San-Ju Lin, Yih-Lan Liu, and Jerry Chih-Yuan Sun. Special thanks to the president of Taiwan E-Learning and Digital Content Association, Chair Professor Ching-Chung Tsai, for his comments and suggestions on the revision of the manuscript. The study was partially supported under the grant number Grant 108-2511-H-009 -013 -MY3 from the Ministry of Science and Technology, Taiwan.

References

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191-215. doi:10.1037/0033-295X.84.2.191

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, *28*(2), 117–148.

Bandura, A. (1997). *Self-efficacy: The Exercise of control*. New York, NY: Freeman.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Multifaceted impact of self-efficacy beliefs on academic functioning. *Child Development*, *67*(3), 1206–1222.

Biggs, J., Hawley, P. H., & Biernat, M. (2018). The Academic conference as a chilly climate for women: Effects of gender representation on experiences of sexism, coping responses, and career intentions. *Sex Roles*, *78*(5–6), 394–408.

Cain, M. K., Zhang, Z., & Yuan, K.-H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior Research Methods*, *49*(5), 1716–1735.

Cheng, Y.-H., Tsai, C.-C., & Liang, J.-C. (2019). Academic hardiness and academic self-efficacy in graduate studies. *Higher Education Research & Development*, *38*(5), 907–921. doi:10.1080/07294360.2019.1612858

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255.

Chou, C., & Lee, Y.-H. (2017). The Moderating effects of internet parenting styles on the relationship between internet parenting behavior, internet expectancy, and internet addiction tendency. *The Asia-Pacific Education Researcher*, *26*(3–4), 137–146. doi:10.1007/s40299-017-0334-5

Chuang, S.-C., Lin, F.-M., & Tsai, C.-C. (2015). An Exploration of the relationship between Internet self-efficacy and sources of Internet self-efficacy among Taiwanese university students. *Computers in Human Behavior*, *48*, 147–155.

Davidson, J. M., & Lyons, M. (2018). Undergraduates as researchers–the impact of active participation in research and conference presentation on psychology undergraduate identity and career aspirations. *Journal of Perspectives in Applied Academic Practice*, *6*(1), 39–46.

De Vries, B., & Pieters, J. (2007). Knowledge sharing at conferences. *Educational Research and Evaluation*, *13*(3), 237–247.

Granziera, H., & Perera, H. N. (2019). Relations among teachers' self-efficacy beliefs, engagement, and work satisfaction: A Social cognitive view. *Contemporary Educational Psychology*, *58*, 75–84.

Greenhow, C., & Gleason, B. (2014). Social scholarship: Reconsidering scholarly practices in the age of social media. *British Journal of Educational Technology*, *45*(3), 392–402.

Greenhow, C., Li, J., & Mai, M. (2019). From tweeting to meeting: Expansive professional learning and the academic conference backchannel. *British Journal of Educational Technology*, *50*(4), 1656–1672.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). New York, NY: Prentice Hall.

Hall, N. (2015). Delineating the learning process in generating a research culture among undergraduate social work students: A Case study of student participation in an academic conference. *Social Work Education*, *34*(7), 829–845.

Harter, S. (1982). The perceived competence scale for children. *Child Development, 53*(1), 87–97. doi: 10.2307/1129640

Henderson, E. F. (2015). Academic conferences: Representative and resistant sites for higher education research. *Higher Education Research & Development*, *34*(5), 914–925.

Hinsley, A., Sutherland, W. J., & Johnston, A. (2017). Men ask more questions than women at a scientific conference. *PloS One*, *12*(10). doi:10.1371/journal.pone.0185534

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. doi:10.1080/10705519909540118

Jacobs, N., & McFarlane, A. (2005). Conferences as learning communities: Some early lessons in using 'back-channel'technologies at an academic conference–distributed intelligence or divided attention? *Journal of Computer Assisted Learning*, *21*(5), 317–329.

Jardine, B., & Romaniuk, J. (2009). *Influences on audience inheritance in Australian television viewing*. Melbourne, Australia: ANZMAC.

Jardine, B., Romaniuk, J., Dawes, J. G., & Beal, V. (2016). Retaining the primetime television audience. *European Journal of Marketing*, 50(7/8), 1290-1307. doi:10.1108/EJM-03-2015-0137

Jokisch, M. R., Schmidt, L. I., Doh, M., Marquard, M., & Wahl, H.-W. (2020). The Role of internet self-efficacy, innovativeness and technology avoidance in breadth of internet use: Comparing older technology experts and non-experts. *Computers in Human Behavior*, 106408. doi:10.1016/j.chb.2020.106408

Jones, T. M., Fanson, K. V., Lanfear, R., Symonds, M. R., & Higgie, M. (2014). Gender differences in conference presentations: A Consequence of self-selection? *PeerJ*, *2*, e627. doi:10.7717/peerj.627

Kellogg, W. A., Erickson, T., Wolf, T. V., Levy, S., Christensen, J., Sussman, J., & Bennett, W. E. (2006). Leveraging digital backchannels to enhance user experience in electronically mediated communication. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work* (pp. 451–454). New York, NY: ACM.

Kimmons, R., & Veletsianos, G. (2016). Education scholars' evolving uses of twitter as a conference backchannel and social commentary platform. *British Journal of Educational Technology*, *47*(3), 445–464.

Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: The Guilford Press.

Kreijns, K., Kirschner, P. A., Jochems, W., & Van Buuren, H. (2011). Measuring perceived social presence in distributed learning groups. *Education and Information Technologies*, *16*(4), 365–381.

Látková, P., Wu, H.-C. J., & Paulsen, R. D. (2009). Examining social capital in the conference setting: A Case study of college student participants. *SCHOLE: A Journal of Leisure Studies and Recreation Education*, *24*(1), 38–49.

Lee, M.-H., Hsu, C.-Y., & Chang, C.-Y. (2019). Identifying Taiwanese teachers' perceived self-efficacy for Science, Technology, Engineering, and Mathematics (STEM) knowledge. *The Asia-Pacific Education Researcher*, *28*(1), 15–23.

Lee, Y.-H., & Wu, J.-Y. (2013). The Indirect effects of online social entertainment and information seeking activities on reading literacy. *Computers & Education*, *67*, 168–177. doi:10.1016/j.compedu.2013.03.001

Little, C. (2020). Undergraduate research as a student engagement springboard: Exploring the longer-term reported benefits of participation in a research conference. *Educational Research*, 1–17.

Luan, H., Geczy, P., Lai, H., Gobert, J., Yang, S. J. H., Ogata, H., Baltes, J., Guerra, R., Li, P., & Tsai, C.-C. (2020). Challenges and future directions of Big Data and Artificial Intelligence in education. *Frontiers in Psychology*, *11*, 580820.

Macfarlane, B. (2007). Defining and rewarding academic citizenship: The Implications for university promotions policy. *Journal of Higher Education Policy and Management*, *29*(3), 261–273.

Mata, H., Latham, T. P., & Ransome, Y. (2010). Benefits of professional organization membership and participation in national conferences: Considerations for students and new professionals. *Health Promotion Practice*, *11*(4), 450–453.

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods*, *10*(3), 259–284.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543.

Mervosh, S., Lu, & Swales, V. (2020, April 7). Which states and cities have told residents to stay at home. *The New York Times*. Retrieved from https://www.nytimes.com/interactive/2020/us/coronavirus-stay-at-home-order.html

Meyer, D., & Hyndman, R. J. (2006). The Accuracy of television network rating forecasts: The effects of data aggregation and alternative models. *Model Assisted Statistics and Applications*, *1*(3), 147–155.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance* (1st ed.). New York, NY: Routledge.

Moore, C., Fisher, T., & Baber, E. (2016). Virtually unknown: Teacher engagement in an online conference. *ELT Journal*, *70*(2), 200–211.

Richardson, J. C., Maeda, Y., Lv, J., & Caskurlu, S. (2017). Social presence in relation to students' satisfaction and learning in the online environment: A Meta-analysis. *Computers in Human Behavior*, *71*, 402–417.

Rosseel, Y., Jorgensen, T. D., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., Rhemtulla, M., Katsikatsou, M., Barendse, M., & Scharf, F. (2019). *lavaan: Latent variable analysis* (0.6-5) [Computer software]. Retrieved from https://CRAN.R-project.org/package=lavaan

Spilker, M., Prinsen, F., & Kalz, M. (2020). Valuing technology-enhanced academic conferences for continuing professional development. A Systematic literature review. *Professional Development in Education*, *46*(3), 482–499.

Sugimoto, C. R., Work, S., Larivière, V., & Haustein, S. (2017). Scholarly use of social media and altmetrics: A Review of the literature. *Journal of the Association for Information Science and Technology*, *68*(9), 2037–2062.

Taylor, T. (2011). Video conferencing us talking face-to-face: Is video suitable for supportive dialogue? *International Journal of Therapy and Rehabilitation, 18*(7), 392–402.

The R Core Team. (2020). *R: A Language and environment for statistical computing* (3.6.3) [Computer software]. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.207.1436

Tschannen-Moran, M., & Hoy, A. W. (2007). The Differential antecedents of self-efficacy beliefs of novice and experienced teachers. *Teaching and Teacher Education*, *23*(6), 944–956.

Wu, J.-Y. (2014). Gender differences in online reading engagement, metacognitive strategies, navigation skills and reading literacy. *Journal of Computer Assisted Learning*, *30*(3), 252–271. doi:10.1111/jcal.12054

Wu, J.-Y. (2017). The Indirect relationship of media multitasking self-efficacy on learning performance within the personal learning environment: Implications from the mechanism of perceived attention problems and self-regulation strategies. *Computers & Education*, *106*, 56–72. doi:10.1016/j.compedu.2016.10.010

Wu, J.-Y. (2020). Learning analytics on structured and unstructured heterogeneous data sources: Perspectives from procrastination, help-seeking, and Machine-Learning defined cognitive engagement. *Computers & Education*, 104066. doi:10.1016/j.compedu.2020.104066

Wu, J.-Y., & Cheng, T. (2019). Who is better adapted in learning online within the personal learning environment? Relating gender differences in cognitive attention networks to digital distraction. *Computers & Education*, *128*, 312–329.

Wu, J.-Y., Hsiao, Y.-C., & Nian, M.-W. (2020). Using supervised machine learning on large-scale online forums to classify course-related Facebook messages in predicting learning achievement within the personal learning environment. *Interactive Learning Environments, 28*(1), 65-80. doi: 10.1080/10494820.2018.1515085

Wu, J.-Y., & Hughes, J. N. (2015). Teacher network of relationships inventory: Measurement invariance of academically at-risk students across ages 6 to 15. *School Psychology Quarterly*, *30*(1), 23–36. doi:10.1037/spq0000063

Wu, J.-Y., & Kwok, O. (2012). Using structural equation modeling to analyze complex survey data: A Comparison between design-based single-level and model-based multi-level approaches. *Structural Equation Modeling - A Multidisciplinary Journal*, *19*(1), 16–35. doi:10.1080/10705511.2012.634703

Wu, J.-Y., Kwok, O., & Willson, V. L. (2014). Using design-based latent growth curve modeling with cluster-level predictor to address dependency. *The Journal of Experimental Education*, *82*(4), 431–454. doi:10.1080/00220973.2013.876226

Wu, J.-Y., Lee, Y.-H., & Lin, J. J. H. (2018). Using iMCFA to perform the CFA, multilevel CFA, and maximum model for analyzing complex survey data. *Frontiers in Psychology*, *9*. doi:10.3389/fpsyg.2018.00251

Wu, J.-Y., Lin, J. J. H., Nian, M.-W., & Hsiao, Y.-C. (2017). A Solution to modeling multilevel confirmatory factor analysis with data obtained from complex survey sampling to avoid conflated parameter estimates. *Frontiers in Psychology*, *8*, 1464.

Wu, J.-Y., & Nian, M.-W. (2021). The Dynamics of an online learning community in a hybrid statistics classroom over time: Implications for the question-oriented problem-solving course design with the social network analysis approach. *Computers & Educatio*n, 104120. doi:10.1016/j.compedu.2020.104120

Wu, J.-Y., & Peng, Y.-C. (2017). The Modality effect on reading literacy: Perspectives from students' online reading habits, cognitive and metacognitive strategies, and web navigation skills across regions. *Interactive Learning Environments*, *25*(7), 859–876. doi:10.1080/10494820.2016.1224251

Wu, J.-Y., Yang, C.C.Y., Liao, C.-H., & Nian, M.-W. (2021). Analytics 2.0 for precision education: An Integrative theoretical framework of the human and machine symbiotic learning. *Educational Technology & Society, 24*(1), 267-279.

Supplementary material

Table S.1. The Virtual Conference Mastery Experience (VCME) Scale

Mastery Experience (VCME)	
S1	I am satisfied with my experience in participating in the keynote speeches of this virtual conference
S2	I am satisfied with my experience in participating in the oral presentations of this virtual conference
S3	I am satisfied with my overall experience in this virtual conference

Table S.2. The Virtual Conference Self Competence (VCSC) Scale

Social Interaction Competence (VCSC-SIC)	
SIC1	I am certain I can make new friends in the virtual conference.
SIC2	I am able to meet my research fellows in the virtual conference.
SIC3	It is easy for me build up connections with academic peers in the virtual conference.
ICT Use Competence (VCSC-ICT)	
ICT1	I am able to switch between rooms and attend more parallel sessions in the virtual conference.
ICT2	I can have more time to prepare my presentations or listen to others in the virtual conference.
ICT3	I am certain that I can save the travel cost and time for attending the virtual conference
ICT4	I am confident that I can switch between slides and other media during my presentation.
Academic Competence (VCSC-AC)	
AC1	I believe I can obtain the latest research development or trends in the virtual conference.
AC2	I think I can concentrate on the presentation contents in the virtual conference.
AC3	I can share my academic works effectively in the virtual conference.

Challenges in Organizing Online Conferences: Lessons of the COVID-19 Era

Margus Pedaste* and Meriliis Kasemets
Institute of Education, University of Tartu, Estonia // margus.pedaste@ut.ee // meriliis.kasemets@ut.ee
*Corresponding author

(Submitted October 16, 2020; Revised November 13, 2020; Accepted December 13, 2020)

ABSTRACT: Travel restrictions regarding COVID-19 have created new challenges for organizing international scientific conferences. Most of the conferences have to be moved to a fully online format. In our study, we analyzed what challenges it created in organizing the International Conference on Advanced Learning Technologies and Technology-Enhanced Learning in July 2020 in Estonia at the University of Tartu. The conference had 131 attendees from all over the world – this resulted in significant challenges due to differences in time zones and difficulties in engaging and socializing people in an online event. In our study, we collected feedback on conference-organization related challenges from five local organizing team members with different roles in the team. Their interviews were analyzed using an abductive approach. The results showed that the challenges could be identified through eight main categories: value, management, timing, program, people, protection, scaffolds, and money. In each of them, several sub-categories were specified. It was concluded that there were both advantages and challenges in organizing an online conference compared to a regular one. For example, fewer challenges are related to travel, accommodation, food and drinks, but more attention needs to be paid to supporting the socialization of people, especially those living in different time zones. Another major challenge appeared to be uncertainty related to the conference budget. Significant advantages were that the carbon footprint of the conference was smaller, the conference was more accessible, and it was easier to solve all the technical issues of the participants.

Keywords: Online conference, COVID-19, ICALT, Zoom, Remo, Abductive analysis

1. Introduction

Numerous guidelines on organizing conferences have been available for a long time. For example, a handbook published by the US General Secretariat Department of Conferences and Meetings Management (Department of Conferences and Meetings Management, n.d.) has specified many processes that could be systematically taken into account when preparing for a conference. For example, management topics as making an agreement with the organizers, assigning coordinators, and forming the organizing committee. Another set of process are concerned with program, e.g., specifics of plenary session. However, often many guidelines seem to be specific for a face-to-face conference, e.g., rooms, a lounge for delegates. Indeed, some of the categories of the process seem to be relevant in whatever format, e.g., establishing communication channels and support for the participants of the conference. Thus, organizing a conference is definitely a big challenge, requiring a strong team to succeed. Therefore, the aforementioned guidelines identified 12 different staff roles and responsibilities in managing and coordinating conferences. Each of them is assigned tasks to complete before the conference, during the conference and after the conference. Finally, the handbook also provides a meetings check-list consisting of 183 items. However, these are rather technical and do not focus on the inclusion and active involvement of the attendees of the conference.

IEEE as the world’s largest professional organization has developed their own guidelines to support hundreds of conference organizers yearly, and their focus has been much more on the inclusion of the attendees. Their manifest states that the conference attendees should feel welcome and included in a conference and that a safe and positive environment encourages attendee participation, fosters collaboration, and builds community (Institute of Electrical and Electronics Engineers (IEEE), n.d.). As a result, they have listed specific guidelines on promoting inclusion, ensuring accessibility for all attendees, and childcare at conferences; as well as ethical guidelines and policies, but also guidelines on more technical aspects such as event emergency management, event conduct and safety statement, and financial transparency in managing expenses and travels.

When synthesizing these guidelines, we can conclude that in organizing a conference, we could systematically focus on at least 20 different categories of activities in each phase of conference organization – before, during and after the conference. In Table 1, we present a few examples of questions in each of these categories in the first phase, *before the conference*. However analogous questions could be specified or even similar follow-up questions formulated for the *during the conference* and *after the conference* phases. For example, before the conference, we need to specify in the venue category what country would be attractive for the attendees, whereas during the conference, we need to ensure that the uniqueness of the country is highlighted systematically, and after the conference, we need to analyze what characteristics of the venue increased/decreased the attractiveness of the conference venue in order to make more informed decisions in the future. Therefore, the same guiding questions could be at least partially used in all three phases of organizing a conference.

Table 1. Examples of questions in organizing a conference (presented are only questions asked before the conference)

Categories of activities	Before the conference
1. Venue	What country would be attractive for the participants? In which city to organize the conference? Is it better to have the conference in the rooms of the host institution or in a conference center?
2. Funding and budget	What has been the budget of similar conferences? What could be the participation fee? Who could be the sponsors of the conference?
3. Timing	What are the best dates? When to start and finish each day? When to have keynotes?
4. Program	How many conference days are needed? What are the presentation formats? What should be the social program?
5. Organizing team	What roles need to be covered? How to ensure good communication in the team? How to secure all the roles?
6. Conference rules	Who has access to the conference materials? What consents need to be asked from the participants? Is taking and sharing photos and videos allowed?
7. Benefit for the organizers	What are the benefits for the country where the conference is organized? What are the benefits for the host institution? What are the benefits for the organizing team?
8. Language and editing	What is the language of the conference? How to ensure good editing of all conference materials? How to support authors in editing their articles?
9. Invited participants	Who should be invited from the organization affiliated to the conference? What benefits need to be provided to the keynotes? How many invited participants to have?
10. Travel and accommodation	What guidelines would be expected by the attendees? What discounts are possible? How to support participants in getting visas?
11. Food and drinks	What meals are provided to the attendees? How to ensure appropriate food and drinks to attendees with different cultural backgrounds? How to ensure enough options to provide food to attendees with different allergies and diets?
12. Rooms	How many rooms are needed in parallel? How many seats are needed in different rooms? What needs to be the setup of the rooms?
13. Technical support	How to set up communication channels to be used already before the conference? What are the technical specifications needed in the conference rooms? How to provide technical helpdesk to the attendees?
14. Communication	What different channels are needed? What needs to be communicated? How to ensure an appropriate amount of information?
15. Website	Where to host the conference website? What information is needed on the website? How to keep the website up to date?
16. Consents	What copyright agreements are needed to publish conference materials?

	What permissions need to be asked to take photos and record videos of the conference? What general regulations of data and privacy protection need to be followed?
17. Social program	What has been organized in the previous conferences? What are the specifics of the host country? How to ensure the inclusion of most of the attendees?
18. Conference materials	What information needs to be presented on badges? What should be given to the attendees in the conference materials? What banners and signs should be used in the conference venue?
19. Sessions	What equipment and supplies are needed in sessions? How to support session chairs? How to ensure technical readiness of rooms?
20. Security and safety	How to ensure a safe physical and virtual environment for attendees? How to protect data of the participants? How to solve emergencies?

The COVID-19 pandemic has started a new era in organizing conferences, although some experiments with innovations have been done earlier as well (see Abbott, 2019). Many conferences have been cancelled or postponed (Viglione, 2020). In some other cases, the conferences have been organized in a fully online format or in a hybrid format, where a few speakers or, sometimes, some more local participants have been invited to the studio of the host but most of the attendees have indeed been participating through online environments (see Bhargava, Farabi, Rathod, & Singh, 2020). Despite the change of the format, the participants still expect to present their latest work and network with their peers (Sarabipour et al., 2020).

We assume that the same questions – and in the same categories – that normally apply in organizing a regular conference also apply in organizing an online conference. However, recent studies have already revealed some notable differences in responding to these questions in different settings. For example, Reshef, Aharonovich, Armani, Gigan, Grange, Kats and Sapienza (2020) analyzed the tips in organizing a conference with 1,100 researchers participating remotely. Their conference was organized in January 2020, before the COVID-19 lockdown period. Therefore, they focused on the idea that the traditional conference format requires change in order to better respond to the academic communities' needs derived from changes in their work-life balance as well as to the technological advances that allow fast and reliable internet connection for teleconferencing. We may also add that online conferences reduce the need for traveling, therefore reducing the conference's environmental impact. Thus, the change of the paradigm in organizing conferences is not only the result of the COVID-19 era but necessary anyway as a response to many changes in the world.

In their analysis, Reshef et al. (2020) concluded that online conferences bring clear benefits for research communities: eliminating the cost of holding in-person events, freedom offered by the use of technologies, academic meetings reaching a wider audience, reduced travel and conference attendance costs, but also a decreased impact on the families of the attendees. However, according to our knowledge, there are no systematic studies on organizing conferences in the COVID-19 era. Valuable studies on understanding the effect and challenges of virtual conferences have been made by analyzing conferences organized before the spread of COVID-19 (e.g., Veldhuizen, Slingerland, Barredo, & Giller, 2020), but in these cases, it was not possible to take into account the authentic situation we have since March 2020. Therefore, we aimed to study how COVID-19 has affected the organization of conferences in all 20 categories of questions specified in the guidelines on organizing regular face-to-face conferences. We believe that this analysis might be valuable for conference organizers, but at the same time, we understand its limitations, as every conference is unique and new lessons can be learned from each new conference.

2. Methods

In this study, we focus on the experience gained in organizing the International Conference on Advanced Learning Technologies and Technology-Enhanced Learning (ICALT, http://icalt2020.ut.ee). This ICALT conference is affiliated to the IEEE Computer Society and IEEE Technical Committee on Learning Technology. The conference has 20 years of history, and in 2020, it was organized as a fully online conference at the University of Tartu in Estonia. We interviewed five key members of the conference organizing team to understand the challenges and solutions in all three phases of the organization process: before the conference, during the conference and after the conference. The collected data were analyzed using abductive content analysis to develop generalizable guidelines on organizing online conferences in the future.

2.1. Context of the ICALT conference

The ICALT 2020 aimed to bring together people who are working on the design, development, use and evaluation of technologies that will be the foundation of the next generation of e-learning systems and technology-enhanced learning environments. The conference has been organized in 20 years in many different countries. Usually, the conference attracts between 100 and 200 participants and between 200 and 300 paper submissions. About 25% of the submitted papers are accepted as full papers, making the competition quite strong. All papers are indexed in IEEE Explore digital database and indexed by several databases, including Web of Science. In 2020, the conference received 171 paper submissions from 40 countries. More authors were from China (50), Taiwan (48), and Brazil (41). Submission numbers are usually higher from the host countries: e.g., there were 33 authors from Estonia in 2020 compared to 9 and 10 in the preceding years. In 2020, a total of 33 papers were accepted as full papers (19.3% acceptance rate), 54 papers as short papers, 26 as discussion papers, and 6 papers were selected for the Doctoral Consortium. The online conference was accessed by 131 attendees.

In 2020, the conference was held from July 6 to July 9 as a fully online conference. The conference was accessible via a website where the registered participants received a password to access the restricted site. The conference proceedings and the conference program were available for downloading. The electronic program of the conference was adapted to the user's computer clock. The schedule was accommodated, as much as possible, to the time zones of the presenters in different sessions to avoid the need to give a presentation during the night. Each session in the schedule had a unique link to an online conference room in Zoom (see http://zoom.us). Several professional accounts were used to enable recording of parallel sessions in the cloud (see Figure 1). The recordings were made available in the program immediately after the end of the session so that attendees could watch them for one month. Social events of the conference (e.g., quiz about Estonia using Kahoot, guided city tour of Tartu, live concert of a local artist) were usually organized as synchronous activities in Zoom. A few local folk dance classes were provided to the participants as recordings that were played during some breaks. The Remo conference tool (https://remo.co) was used to facilitate discussions in small groups during the breaks (see Figure 2). This tool allows users to autonomously form small discussion groups around virtual round tables, share their screens and use a whiteboard for their discussions.

Figure 1. Screenshot of the opening ceremony of the ICALT2020 conference (screenshot published with the permission of the persons in the screenshot)

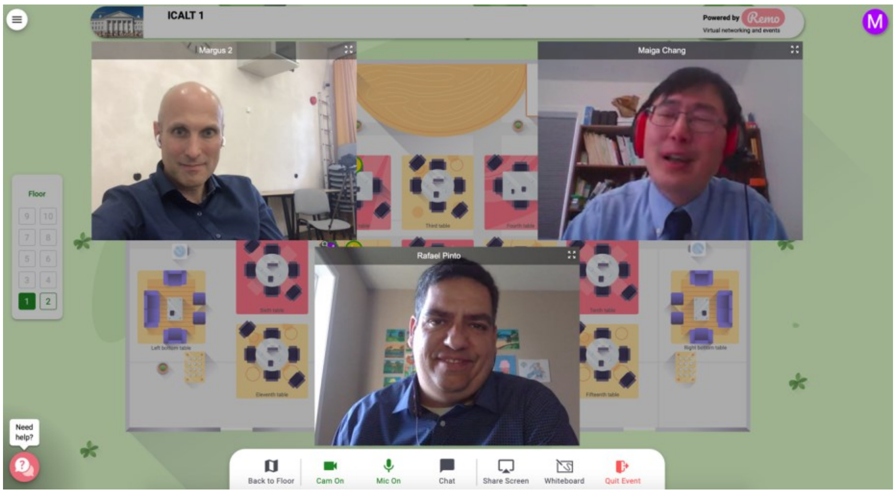


Figure 2. Screenshot of the Remo virtual lobby room of the ICALT2020 conference (screenshot published with the permission of the persons in the screenshot)

2.2. Context of Estonia

Organizing an online conference in Estonia might be easier than in some other countries. Internet connection is considerably good almost everywhere, and the country's schools and universities have designed their learning systems to benefit from the well-developed digital infrastructure. Significant improvements have also been made in e-learning, which provides a good basis for adapting conferences to online format. The lecture halls of the universities are usually well equipped with technology to have online conferences. The Institute of Education at the University of Tartu, the host of the ICALT2020 conference, has provided a mostly online international master's program on Educational Technology since 2017. After testing several digital tools, the staff opted for Zoom as the principal tool to be used in the learning process, as it proved to be a reliable environment for online sessions even with poor internet connection. In addition, Zoom allows sharing of screens, optimizing the video broadcast, recording of the sessions to the cloud, and dividing students in groups working in breakout rooms.

2.3. Data collection and analysis

The data for the current study were collected by five semi-structured interviews using questions about the three phases in the conference organization procedure. The interviews were conducted three months after the conference. Two questions were asked about the preparation phase: (1) "What issues/working items should be taken into consideration?" and (2) "What challenges/obstacles were faced during the preparation phase?" Three questions were asked about the implementation phase: (1) "What issues/working items should be taken into consideration?"; (2) "What challenges/obstacles were faced during the implementation phase?"; and (3) "How did you resolve those difficulties? Please share the lessons learned and provide suggestions." Two questions focused on the evaluation and feedback phase: (1) "How do you evaluate the success of the conference?" and (2) "What do you think of participants' feedback?" Three of the interviews were done by one of the authors of this article and two were conducted as self-reflections of both authors of this article.

All five respondents had key roles in organizing the ICALT2020 conference. The roles covered by the five selected persons were the following: conference general chair, local committee chair, publicity chair, financial chair, conference manager, and coordinator of technical support. The experience these people had regarding organizing and participating in different conferences varied a lot. Most of the team (except for the conference manager) had recent experience of organizing the in-person EAPRIL2019 conference (https://www.eapril.org/eapril-2019). This conference had 464 participants from 33 different countries, mainly from Europe. Only one respondent (the one in the roles of the conference general chair and local committee chair) had any experience of participating in previous ICALT conferences (in five different conferences in 2008, 2009, 2017, 2018, 2019). This person had also chaired the EAPRIL2019 conference and acted as a member of the organizing committees of many other international conferences held in different countries. In addition, he had participated in more than 100 academic conferences over 20 years and acted as a senior member of IEEE (which is affiliated to the ICALT2020 conference) and vice-chair of the IEEE Estonia section, while the experience of the other respondents was significantly more limited.

The five interviews yielded 80 idea units containing a total of 2904 words (an average 36 words per idea unit). The number of idea units in an interview ranged from 6 to 32. The highest number of idea units came from the most experienced respondent; the number of idea units from other respondents varied from 6 to 17.

The interview data were analyzed using abductive content analysis. Abductive analysis (Tavory & Timmermans, 2014) enables to connect theory and empirical data by combining the deductive and inductive process. It is a process of theorizing grounded in pragmatism to make sense of the data. In this case the researcher should have a very good overview of theories on the research topic and then he/she needs to move recursively back and forth between observations and theory. In our case the "theory" was an extensive experience of the article's first author's participation and organization of international conferences in face-to-face settings. It was extended by the guidelines provided by several associations for organizing conferences. This practical and theoretical knowledge was taken into account in analyzing interviews by thinking on all 20 categories of conference organization activities derived from literature (described in the introduction of the article) and combining this with more inductive content analysis by finding the main themes in the responses of the interviewees and categorizing these into larger units. Thus, the recommendations for the future, as the new "theory" were developed in synthesizing theory and findings. The abductive process was conducted by one researcher and the other author of the paper reviewed the analysis and conclusions based on her experience and knowledge. Thus, a limitation of this method might be replicability of the findings because these depend very much on the researcher conducting the analysis. However, its validity should be increased due to the extent of the theoretical and practical knowledge of the researcher.

3. Findings

We started the analysis of the idea units using the 20 categories identified based on the aforementioned conference organization guidelines (Department of Conferences and Meetings Management, n.d.; IEEE, n.d.). The analysis showed that the respondents focused on almost all categories of activities in their feedback. There were only 3 categories that were not mentioned: language and editing, invited participants, and food and drinks. These could be considered topics that caused fewer challenges in organizing an online conference. However, this could also indicate that there were no particular memories regarding these categories. Each respondent could provide more than one idea unit in each category; in a few cases, one idea unit covered two categories. The most frequently mentioned categories were *venue* and *funding and budget* (both 9 times), followed by *sessions* (8), *social program* and *security and safety* (6), *organizing team*, *technical support*, *communication* (5), *timing*, *travel and accommodation* (4), *program*, *benefit for the organizers* (3), *website*, *consents* (2), and *conference rules*, *rooms*, *conference materials* (1). The main categories found were (1) management, (2) protection, (3) timing, (4), people, (5) scaffolds, and (6) money (see Figure 3).

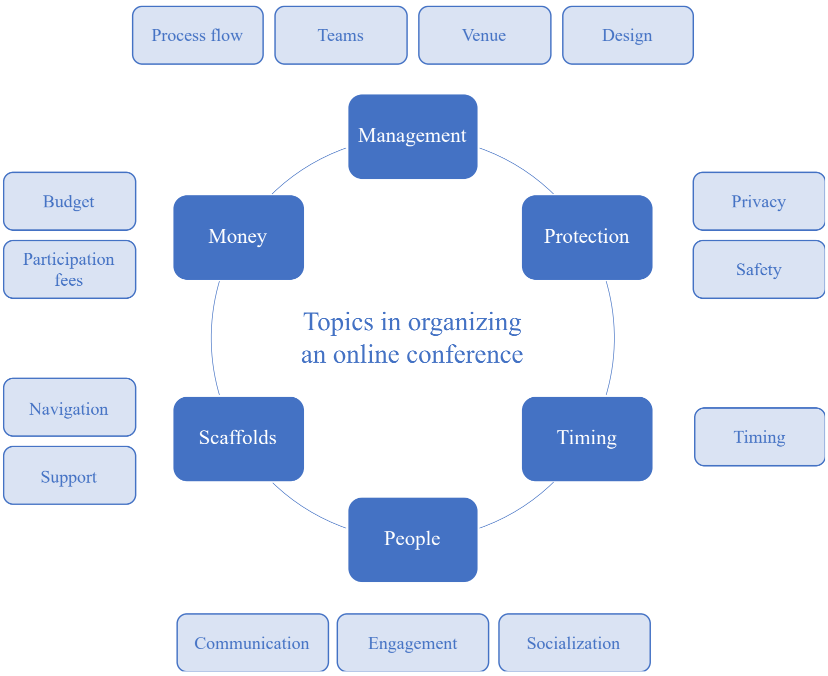


Figure 3. Main categories and sub-categories of topics of attention in organizing an online conference

3.1. Management

The first general theme specified in the analysis was *Management*. It was found that the organization of a conference should start with designing a clear process flow. The respondents indicated that it would have been very useful if that had been provided by IEEE as a professional organization affiliating hundreds of conferences every year. If the flow is clear then it is easier to adapt a regular conference for an online format and to plan activities that should be completed as an input for other people in completing their tasks. The tasks according to the process flow need to be assigned to different teams were several people can support each other or take over the roles if needed.

One of the tasks for the teams according to the flow is the management of the venue. In case of online conferences it consisted of topics as physical headquarters, online tools for sessions and social events, the technical setup of the online rooms. The challenges regarding the conference *venue* were distinctly different compared to a regular conference. Something was mentioned by all respondents. The main challenge was finding an online tool that would make it possible to provide all the services of a regular conference in an online format. Another challenge was time – there was not enough time to search for the best tools and test them all. Therefore, we had to be very focused. Third, the conference team also searched for professional conference organizers, but their solutions had some limitations in spite of high cost. Therefore, in the end, we had to select a set of different tools based on the existing experience. The conference manager noted the following:

Functionality was our first consideration, so that everything that is part of a conference could also be part of the conference in the online version. The fewer different apps a participant has to use, the better. We decided in favor of Zoom, where all sessions and social events took place.

The fourth topic categorized under management was design. It mainly concerned the design of the conference website so that all relevant information about the process flow, teams and venue is also available for the potential participants clearly and attractively. In our case, it was found that the website was very functional but with a simple design that could have been improved. Again, it was mentioned that it would have been good if, for a conference organized under the umbrella of a large organization as IEEE or a conference that has been organized already 19 times before, it would have been good to design a website that could be reused from year to year with small improvements made as needed.

We used a password-protected website, where participants were able to access all sessions in Zoom using links in an electronic agenda. Additionally, Remo was used as a tool for small group discussions during breaks between parallel sessions. However, it would have been best if the tools had been integrated into one system. In addition, we set up physical conference headquarters in one of the university's buildings and offered the local presenters an additional opportunity to give their presentation at the headquarters. However, this option was only used by a few presenters.

3.2. Protection

Protection-related topics focused on two sub-categories. Main challenges were about privacy but a few comments were also made on safety. The challenge was to get informed consent from all participants of the online conference to record the sessions to make them rewatchable for the registered participants following the General Data Protection Regulation (GDPR). We prepared for blurring parts of videos in case someone would not agree to be recorded. However, it was not needed in the end. We simply stopped recording the session when a person was presenting who did not agree to be video recorded. There were only a few of such persons. When this person participated in sessions where others were presenting, we asked him/her to close the camera. We also prepared for security attacks in the sessions (e.g., each session had a technical assistant always present in the role of host with the rights to remove any participant from the session if needed), but this time these plans were not needed. No issues were reported. A minor issue that we encountered was sharing screenshots taken in the sessions using personal social media accounts, even though we had informed each participant that making or sharing any recordings of the conference sessions was not allowed.

Safety was mentioned less but meant more different aspects. First, it was noted that participants felt themselves safe because of the technical support available in the sessions. Second, some presenters had concerns about hacking the sessions by some unwanted people who may start to attack the people with inappropriate comments or content. Titipat, Tulakan, Isil, Wyble and Kording (2020) has noted this behavior and named it "Zoombombing." Therefore, we instructed technical support on how to deal with these issues (e.g., dropping out these people and blocking them to of re-joining) but we did not have such cases. Third, we had to ensure that the conference is only for these people who have paid the conference fee or who are granted with free complimentary access (e.g., the conference chairs, track chairs or local organizing committee members). Therefore, we disseminated login information only shortly before the start of the conference and asked everyone to identify themselves at the first login. Later, the technical support also checked in the sessions if there are only the people who have checked in. Of course better solution might have been to create individual access accounts for everyone, but the webpage solution we used did not have such feature and we did not have time to start to build up a webpage from scratch.

3.3. Timing

The timing was challenging because of the different time zones of the participants in an online conference. Therefore, it was very important to adapt the program to the presenters' time zones and to make the recordings of the sessions available as soon as possible after each session. The technical support saw that the number of participants in synchronous sessions was about the same as the number of people who watched the recording. Thus, the idea to make the recordings available for the ones in very different time zones worked well. In general, we decided to start the conference day in the morning according to Greenwich Mean Time and to finish in the evening so that the local committee did not have to work during nights. Indeed, the conference days were very long for the local organizers.

3.4. People

The main category *People* had more notes than any other category. First, it was challenging to communicate the conference to the target group. In the end, we had quite many participants and according to our interpretation, there are two main reasons of this: (1) the ICALT community is well established so that many people have attended many previous conferences and are "loyal" to the conference, (2) the accepted papers are published in the conference proceedings and indexed in databases only if at least one author of a paper registers to the conference and presents the paper in the conference.

The two other sub-categories were even more related to the conference days – engagement of participants in the main sessions and socialization during the breaks and in social events. In an online conference, both are challenging. In the sessions, we allowed participants freely to start discussions in the chat and we also allowed to use their microphones to have audio-discussion. More challenging was supporting the follow-up discussions during the breaks. It turned out that it is not a good solution if people need to move between different apps. The social program seemed to be appropriate (not too heavy or with too few events) but we learned that it need be advertised more in the scientific program or even by the sessions chairs at the end of their sessions. First, we planned to use Remo for follow-up discussions and open the Zoom rooms only a few minutes before the sessions and close them immediately after the sessions to allow technical support to update links to recordings in the schedule and prepare for the next session. However, during the conference, it turned out that Remo was not convenient, because it was a different tool. It was much easier to come to Zoom a few minutes before the start of the session and stay there a bit longer to have an additional discussion. Thus, we revised the plan. The coordinator of technical support concluded:

Questions were asked at the end of the session, but there was also a small period of time when the session had not started yet and participants were chatting as well. At first, we thought that all this kind of social chat and discussion would happen in Remo, but it did not work out as planned. Also, some participants seemed to know each other and they connected better. Remo failed because you have to go there separately, Zoom was like a common corridor between the sessions' "rooms"…

In the *social program* we expected more active participation. It turned out to be not as important as in regular conferences. One explanation for this could be that often, people did not see each other's images, did not see the emotional reactions, and this decreased active communication and participation. The conference manager found that the aim of the social events was not to "imitate" the events of a real conference but to provide something that should work in an online event:

With the live concert of Puuluup [a local artist] we wanted to do something that would be Estonian enough, but with a twist. The quiz and the video tour of the city of Tartu were to introduce our university city. Comedy Night was also Estonian-themed and, in my opinion, relatively successful.

The other respondents also found that those who participated in social events (usually around 20 people) were satisfied and engaged. Thus, it might also be the case that people are not yet used to participating in the social events of online conferences. Therefore, these need to be advertised more in the case of future events.

3.5. Scaffolds

Next, we identified two types of *Scaffolds* of an online conference. First, there was available a built-in scaffold that was named navigation – the web-site, electronic program and guidelines had to make navigation during the conference as simple and intuitive as possible. In general, it worked out well. The main issue was that people did not use actively Remo tool that was meant to enable small-group discussions between presentations. Second, there were many notes about technical support. It turned out that the support provided by technical assistance in every session was very much needed. There were no complaints regarding this. It has been noted in other studies as well that in an online format it is easier for moderators and technical support to control the flow of the discussions (see Price, 2020). The general helpdesk was even almost not needed – email turned out to be enough to cover the more specific needs. However, it was found that the support of the organization affiliating the conference might be stronger, e.g., the set of tools, template for a website.

Technical support seemed to be sufficient. In each parallel session, there was a technical assistant who was often approached with many different questions, not only with the technical ones – this role appeared to be very important. As technical support was attending every Zoom session then they were easily reachable and visible all the time. The same people formed a team who planned all the technical details of the conference. There we also prepared guidelines describing the conference flow and use of different tools, e.g., how to make a presentation or participate in a session. In addition, we provided the helpdesk using different communication channels (email, Zoom and phone) but this was even not needed – very few people contacted this.

Communication appeared to be a bit more difficult. In an international conference there had been used in the past some communication channels in case of what it was not clear who has access to the accounts and who are the target groups. It caused a bit of confusion. Also, it was not always clear what information needs to be communicated, when it need to be communicated, who provides this information and who has to submit it to a different channel. Even the local organizing team was sometimes in lack of information. The publicity chair found:

People cannot be flooded with information. Of course, you never know what is enough for whom, what is too much for someone. Due to the change in format, the timeline was very tight and therefore sharing of information was sometimes left to the last minute. A lot of effort was later put in preparing the certificates to the participants.

3.6. Money

Finally, there were also two categories of *Money*-related comments. A challenge was caused by the uncertainty in the registrations. It was not possible to estimate the number of participants and plan the costs without knowing the income. Also, it was not possible to reduce the conference fee without a revised budget although it was very clearly expected by the participants. Thus, we had to take risks and be conservative in budgeting and the end we did not use all of the income.

The conference budget was first prepared for a regular conference. In March 2020, we decided to wait a bit to see how the developments regarding COVID-19 unfold. In April, we had to develop different scenarios for the conference, all of which also affected the budget: (1) the conference will be held as planned, but the number of participants will be smaller than expected because people are afraid of travelling; (2) the conference will be organized in a fully online format; (3) the conference will be organized in a mostly online format; (4) the conference will be canceled; (5) the conference will be postponed. By that time, the submission deadline had already passed, and even the results of the reviews had been announced to the authors. Thus, the registration period had to start and some of the participants had already registered according to the original conference fees. Finally, we decided that considering the situation, we would have to organize the conference in a fully online format.

We anticipated that the change of the conference format might have a significant effect on the potential income, but also costs. First, it turned out that potential sponsors were not interested in supporting an online conference. For example, Enterprise Estonia and Tartu City Government support conferences if the participants stay in local hotels for a certain number of nights (at least 300 or 200 nights per conference, respectively). This reflects the idea that they support local conferences if the participants spend enough money in Estonia by buying the services and goods of Estonian companies. This is not the case at online conferences; therefore, they are not interested in sponsorship.

The benefit of the online format of the conference was that we did not need such a large budget anymore for conference rooms, meals and hosting keynote speakers and other international guests. Therefore, we tried to define a reduced fee for participants, but it was difficult, because it was not possible to estimate the number of paying participants based on the experience of previous years. Indeed, we finally decided to reduce the author fees by about 30% and introduced a completely revised very low fee for non-presenting authors. However, it took the financial manager more time than expected to reorganize everything that had already been planned: to cancel preliminary agreements, pay back part of the already paid fees and communicate with participants that you never meet in person. Also, some issues rose from the fact that IEEE expects the budget in US dollars, but in Estonia, we had to calculate everything in euros. Thus, the financial manager concluded three months after the conference:

We had all the rooms planned ... It took extra hours to rearrange the original plans. Because of COVID-19, people had to do some unexpected additional tasks, it was not pleasant but had to be done to adapt to this new situation. The money had to be reimbursed and it took some effort to get all the required correct bank details … It continues to this day, I am currently putting the last invoice on track. Due to the change in exchange rates between the euro and US dollar, we need to transfer money to IEEE. In addition, more documents were expected due to coronavirus … Every little thing took more time than initially planned. Communication with both banks and our university accountants…

Finally, however, it turned out that some profit had been made that had to be paid to the IEEE and could be used for the next ICALT conferences. This was because we had underestimated the number of registered participants. We were more conservative than needed. We assume that the high number of paying participants was the result of the conference policy. The conference general chair explained:

Budget-wise it was important that the conference registration was bound to publishing the article in the conference proceedings; otherwise it will not be indexed in the databases (it has always been so, but the risk of decreased number of participants is higher for online events, where there seem to be fewer other benefits for the participants).

4. Discussion

In combining the findings from our abductive analysis and guidelines provided earlier for organizing face-to-face conferences revealed eight main categories of activities of organizing a conference (see Table 2). The guidelines for face-to-face conferences indicated two categories that were not highlighted by our interviewees. First, in online conference management, there were not indicated challenges regarding travel and accommodation or food and drinks. This is good in many aspects. For example, Titipat et al. (2020) have indicated several disadvantages of the regular conference, e.g., they have a massive carbon footprint, they are time-consuming, and there are high costs involved in attending the conference – as a result, these might be not so accessible for early career researchers or participants from country with a lower level of economic development. Besides, there might be other reasons why some people do not want to attend a regular conference but are register to a virtual one (see Price, 2020). Thus, the shift of conference from offline to online might have significant value-related pros as well.

Second, the value of the conference for the organizers was not highlighted in the interviews. The reason for this might be that the program was not compiled by the interviewed local committee members and the value of the conference was not the topic to discuss in this team. The value has been specified through the years the conference has been organized within the ICALT community. The value for the local team is more evident for these people who are active members of the academic international community of this particular conference and only the conference chair / local committee chair belonged to this group.

Indeed, it could be easily argued that the program and the value of the conference are also important topics for any online conference. Thus, in conclusion, we can say that there are eight main categories of topics to take into account in organizing a conference. We found these in analyzing the feedback on an online conference but we believe these could well cover the categories needed in regular conferences as well. Of course, the challenges in each category might be different.

Table 2. Categories of activities of organizing a conference

Sub-categories from the abductive analysis	Main categories	Sub-categories from the analysis of guidelines for organizing face-to-face conferences
Design	Management	Website, conference materials
Process flow		Timing
Teams		Organizing team, invited participants
Venue		Venue, sessions, rooms
		Travel and accommodation
		Food and drinks
Privacy	Protection	Security and safety, consents
Safety		
Timing	Timing	Sessions, timing
Communication	People	Communication, website, conference materials, language and editing
Engagement		Sessions
Socialization		Social program
Navigation	Scaffolds	Conference rules, conference materials
Support		Technical support
Budget	Money	Funding and budget
Participation fees		
	Program	Program
	Value	Benefit for the organizers

By reflecting the experience in organizing ICALT2020 online conference we propose that organization of an online conference should start from discussing the *Value* of the conference to the participants, organizing team, and to the organization affiliating the conference. Next, there needs to be set focus on the *Management*, starting from specifying the process flow, teams, invited participants, venue, website and if needed then also travel and accommodation plus food and drinks. Third, focus should be shifted on the *Program* and related to this to *Timing*. Finally, the people-related topics need be decided on the level of details – how to ensure communication, engagement and socialization of *People*, how to provide *Scaffolds* they need, and how to build *Protection* to have a secure and safe conference. In parallel of all of the other categories is the discussion on *Money*. Many decisions depend on the availability of the budget and the budget depends very much on the value for the participants, the attractiveness of the program, the timing and venue of the conference – even in case of an online event.

The category-based approach on defining the challenges in organizing an online conference might help to structure the challenges and to reduce the risks. This is something that has been not often used. For example, Reshef et al. (2020) list different categories, e.g., the format, the team, the infrastructure, the program, the hubs, the poster session but these are not systematized. Both Reshef et al. (2020) and (Department of Conferences and Meetings Management, n.d.) have distinguished between processes before the conference, during the conference and the latter also the process after the conference. However, this is a general question about the workflow that could be applied in case of all the categories identified in our analysis. For example, communication needs to effective in all these three phases.

Regarding the challenges we also need to agree that sometimes the first impression might be not as clear as looking from distance. Reshef et al. (2020) found in their study that the initial reaction of their community after the event was overwhelmingly positive but more challenges revealed in the post-event surveys. As in case of our study socialization appeared to be one of the major challenges, they found that emotional engagement due to the lack of in-person interaction remained still an open question; e.g., the lack of applause after each talk and stimulation in small-group discussions. "Socialization challenge" is the only generalizable challenge in organizing online conferences based on our study. It has been noted in other studies as well. For example, Salomon and Feldman (2020) found that chat rooms with speakers after the session of during the break would give time for additional questions and debate. In our case we tried to introduce Remo as a special tool to support small-group discussions. However, we failed because it was a separate tool and it was not easy to navigate in the system to use it. In addition, based on the findings of Salomon and Feldman (2020) it might be suggested for the future that the presenters need to come to a specific room after the session for follow-up discussions (or the online room of the session could be kept open for a longer time).

Thus, it seems that the online format supports well one of the main goals of the scientific conference – presenting the work of the participants – but still fails a bit in the other main goal – to provide participants with a platform for networking with their peers (see Sarabipour et al., 2020). The same has been indicated by Abbott (2019) who found that the virtual conferences help to significantly decrease carbon footprint of humans but usually the people still would like to have at least one face-to-face annual conference to forge personal connections and collaboration because of the psychological needs of the humans. However, Titipat et al. (2020) provide for consideration one more idea to support networking. They used in their online conference a matching algorithm to facilitate one-on-one meetings. This could help to find the participants with whom you have a common interest and if these meetings are set up by artificial intelligence behind the data then it could enhance networking. Of course, there might be hidden some new privacy-related issues in applying these algorithms.

One more benefit found in organizing ICALT2020 conference online is that we did not have any issues with visas. In the previous years, it has been one of the major issues and; therefore, we assigned a person in our organizing team to support all attendees with visa-related questions because as it is also noted by Hu (2018) that these issues might hinder scientists to participate due to bureaucracy or even travel bans. According to these aspects, online conference might be more inclusive to provide solutions for flexible accessibility in case of different needs and restrictions.

5. Conclusions

Our analysis of the feedback collected from organizers of the ICALT2020 conference showed that it is indeed possible to organize conferences in COVID-19 era. There are some cons but also pros. We found that in case of an online conference there are not so many challenges with travel and accommodation or food and drinks but significantly more important are the issues of protecting participants' privacy and ensuring a safe online environment – one of the major challenges before the conference and during the conference. One of the biggest challenges during the conference seemed to be to support socialization, especially when the participants live in very different time zones. The program could be adapted to the time zones and the video recordings even increase flexibility in participation but the challenge of connecting people remains. The third remarkable challenge of online conferences is uncertainty in the budget. It is more difficult to estimate the number of participants and to predict the income from participation fees. The technical challenges are important in case of both formats – regular and online – but in case of the online conference, it seemed that it was even easier to handle all these issues immediately. When the participants are online then they can always ask questions from technical support and they could be supported without disturbing the presenter of the session.

The conclusions of our study need to be indeed taken with some cautions. One of our limitations was that we did not use a questionnaire to get evaluations of a larger number of conference participants, including the ones who have experience in organizing previous ICALT conferences. Second, we did not interview all organizing team members, e.g., the international team members responsible in the program. However, we indeed covered the representation of all major roles in the team.

Further studies could focus also on collecting logfile data and data about participants' satisfaction in the conference and with different formats of sessions. Also, it would be valuable to compare different platforms in enabling all processes of conference management, e.g., Crowdcast (see Goodman, Wyble, Achakulvisut, Bilgin, Van den Bossche, & Kording, 2020; Titipat et al., 2020) in contrast to a solution where different tools were combined as in our case Zoom, Remo, online schedule, and website.

Acknowledgement

We sincerely thank all the respondents from the ICALT2020 local organizing committee who kindly participated in our study.

References

Abbott, A. (2019). Low-carbon, virtual science conference tries to recreate social buzz. *Nature*, *577*(7788), 13-13. doi:10.1038/d41586-019-03899-1

Bhargava, S., Farabi, B., Rathod, D., & Singh, A. K. (2020). The Fate of major dermatology conferences and meetings of 2020; are e-conferences and digital learning the future? *Clinical and Experimental Dermatology*. doi:10.1111/ced.14272

Department of Conferences and Meetings Management, General Secretariat (n.d.). *Operational handbook for the organization of meetings and conferences away from headquarters*. Retrieved from http://www.scm.oas.org/Manual/OperationalHandbook.pdf

Goodman, D., Wyble, B., Achakulvisut, T., Bilgin, I., Van den Bossche, S., & Kording, K. (2020). How to Run Big Neuroscience Conferences Online–Neuromatch. io. *Medium*. Retrieved from https://medium.com/@kording/how-to-run-big-neuro-science-conferences-online-neuromatch-io-49c694c7e65d

Hu, J. C. (2018). *This Neuroscientist's poster showed how US travel bans stifle ground-breaking research.* Quartz. Retrieved from https://qz.com/1454743/this-neuroscientists-postershowed-how-us-travel-bans-stifle-groundbreakingresearch

Institute of Electrical and Electronics Engineers (IEEE) (n.d.). *General guidelines. Guidance for developing safe, secure, and inclusive conferences*. Retrieved from https://ieeemce.org/planning-basics/general-guidelines/

Price, M. (2020). Scientists discover upsides of virtual meetings. *Science*, *368*(6490), 457-458. doi:10.1126/science.368.6490.457

Reshef, O., Aharonovich, I., Armani, A. M. Gigan, S., Grange, R., Kats, M. A., & Sapienza, R. (2020). How to organize an online conference. *Nature Reviews Materials*, *5*, 253-256. doi:10.1038/s41578-020-0194-0

Salomon, D., & Feldman, M. F. (2020). The Future of conferences, Today: Are virtual conferences a viable supplement to "live" conferences? *EMBO Reports*, *21,* e50883. doi:10.15252/embr.202050883

Sarabipour, S., Schwessinger, B., Mumoki, F. N., Mwakilili, A. D., Khan, A., Debat, H. J., Sáez, P. J., Seah, S., & & Mestrovic, T. (2020). Evaluating features of scientific conferences: A Call for improvements. *BioRxiv*. doi:10.1101/2020.04.02.022079

Titipat, A., Tulakan, R., Isil, B., Wyble, B., & Kording, K. P. (2020). Improving on legacy conferences by moving online. *eLife*, *9*. doi:10.7554/eLife.57892

Tavory, I., & Timmermans, S. (2014). *Abductive analysis: Theorizing qualitative research*. Chicago, IL: University of Chicago Press.

Veldhuizen, L. J., Slingerland, M., Barredo, L., & Giller, K. E. (2020). Carbon-free conferencing in the age of COVID-19. *Outlook on Agriculture*, *49*(4), 321-329.

Viglione, G. (2020). A Year without conferences? How the coronavirus pandemic could change research. *Nature*, *579*(7799), 327-328. doi:10.1038/d41586-020-00786-y

Yang, S. J. H. (2021). Guest Editorial: Precision Education - A New Challenge for AI in Education. *Educational Technology & Society*, *24* (1), 105–108.

Guest Editorial: Precision Education - A New Challenge for AI in Education

Stephen J. H. Yang
National Central University, Taiwan // Stephen.Yang.Ac@gmail.com

ABSTRACT: As addressed by Stephen Yang in his ICCE 2019 keynote speech (Yang, 2019), precision education is a new challenge when applying artificial intelligence (AI), machine learning, and learning analytics to improve teaching quality and learning performance. The goal of precision education is to identify at-risk students as early as possible and provide timely intervention on the basis of teaching and learning experiences (Lu et al., 2018). Drawing from this main theme of precision education, this special issue advocates an in-depth dialogue between cold technology and warm humanity, in turn offering greater understanding of precision education. For this special issue, thirteen research papers that specialize in precision education, AI, machine learning, and learning analytics to engage in an in-depth research experiences concerning various applications, methods, pedagogical models, and environments were exchanged to achieve better understanding of the application of AI in education.

Keywords: Precision education, Artificial intelligence, Learning analytics, Human-centered AI

1. Introduction: New challenges for AI in education

This special issue concerns the use of artificial intelligence (AI) by focusing on new design methods and tools that can be evaluated and leveraged to advance AI research and educational policies and practices for improving teaching and learning. New technology brings new challenges for the application of AI in education. In addition to technology-oriented intelligent tutoring system and robotics, researchers have now begun paying more attention to education-oriented pedagogy, cognition, and humanity (Luan et al., 2020). With the advancement of deep learning and transfer learning, AI research in education is encountering new challenges in terms of reshaping the research trend from general-purpose intelligence to transfer of intelligence, from computation to cognition, from customization to adaptation, from known to unknown, from technology to humanity, and from the one-size-fits-all approach to precision (Yang, 2019). Researchers who used to focus on cultivating common sense through measures such as utilizing knowledge application for general-purpose intelligence are now moving toward transferring intelligence, which is the transfer of trained knowledge from one domain to another. In particular, the advancement of natural language processing (NLP) has enabled the application of pre-trained knowledge to fine-tune domain. In addition to outperforming in terms of computation ability, AI is now moving toward human-perceived areas, such as perception, emotion, psychology, and cognition, which require cognitive thinking. AI algorithms have progressed from being customized beforehand for use on different people to being adapted to fit individual needs in real time. AI algorithms not only perform reasoning to search for insights but also can uncover the unknown about the unknown, such as hidden values and unknown results. AI can improve human productivity through technology, in addition to promoting human intelligence through machine intelligence. From being used to provide one-size-fits-all type of solutions, AI can now offer a one-of-a-kind precise solution for each individual. To address the challenges faced, the focus of AI research is shifting from technology-oriented focus on increasing production and performance to humanity-oriented focus on augmenting human intelligence with machine intelligence.

2. Precision education: Research goal, topics, and implementation

Precision education was inspired by the precision medicine initiative proposed by the former US President Obama in his 2015 State of the Union address. The emergence of precision medicine revolutionized the one-size-fits-all approach to the treatment of diseases by considering individual differences in people’s genes, environments, and lifestyles and by improving the diagnosis, prediction, treatment, and prevention of diseases. Similar to the medicine field, the design of the current education system does not fully consider students’ IQ, learning styles, learning environments, and learning strategies. Inspired by precision medicine, precision education is an innovative approach that emphasizes the improvement of diagnosis, prediction, and treatment and focuses on preventing at-risk students. In general, at-risk students refer to students who are predicted to show low academic performance or drop/withdraw from a course, or students who demonstrate low engagement in terms of learning behavior, emotion, and cognition.

The research goal of precision education is to identify at-risk students as early as possible and provide them with timely intervention through diagnosis, prediction, treatment, and prevention. The research steps of precision education include conducting a diagnosis of students' engagement, learning patterns, and behavior and making predictions concerning students' learning performance and improving predictive models, followed by devising treatment and prevention plans through timely intervention from teachers, learning strategies, and activities. The research topics of precision education can be categorized into governance and policy and technology and practice. Important topics such as ethics, norms, rules, and other concerns related to precision education fall under governance and policy. The effect of precision education on emerging pedagogical environments, such as MOOCs (massive open online courses), e-books, coding, AR/VR, robotics, and games, warrant exploration, in addition to examining critical factors that influence students' learning performance under the governance of precision education. Furthermore, the influence of teachers' intervention on students' learning performance after the governance of precision education merits investigation. The important topics that fall under technology and practice include the design of pedagogical models and tools, learning strategies and learning activities, and evaluation and assessment methods. Data analytics concerning precision education, including text, audio, image, and video analytics, and data visualization for precision education, including dashboard and simulation, also merit investigation.

Precision education can be implemented by applying AI and learning analytics to identify at-risk students and by providing timely intervention to improve teaching quality and the learning outcomes of students. To improve teaching quality, teachers need to identify students' issues of concern and provide feedback. Identification of areas or topics in which students are struggling is critical for the governance of precision education. These measures can enable teachers to achieve better understanding of how their content is being used and how effective it is, thereby facilitating continual enhancement. For optimal learning outcomes, the educational content needs to be tailored to students' level of understanding and students' performance needs to be monitored so that teachers can accordingly modify their teaching to meet students' needs. To take charge of their own learning, students should be enabled to become aware of their performance in comparison with their peers and continually assess whether they are keeping up with the group. Moreover, students should be trained to identify gaps in their prerequisite knowledge and key study skills in which they are lacking. These measures will facilitate the development of students' skills and knowledge in a more personalized and self-paced manner. Students should be provided with an assessment of their progress and information on measures they should adopt to meet their educational goals. Obtaining students' consent for data collection and use will ensure their privacy.

Precision education can be further improved through smart evaluation. The diagnosis and smart evaluation of students' learning activities can be achieved through pre-class preview, reflection, oral reports, and assignments, teaching of special topics, and essay writing and by introducing the examination mechanism into the smart evaluation system. Non-objective factors such as subjectivity, knowledge category, and personal preference hinder manual evaluation. Therefore, the provision of automatic evaluation, question scoring, and feedback can improve the non-subjective factors concerning manual evaluation. The generation and adoption of new deep learning algorithms in the past few years have facilitated the application of deep learning in natural language processing (Transformer, BERT, and GPT3) and has been shown to be more effective than the previous generation of deep learning (CNN, RNN, and LSTM) and traditional machine learning algorithms (LR, SVM, and RF). These new algorithms can achieve performance that is closer to human behavior. To conduct smart evaluation, teachers provide textbooks (textbooks and slides), and the system employs natural language processing technology for text summarization and interception of key concepts in the text content (textbook). Furthermore, the system automatically generates test questions and reference answers (automatic question generation) from the passage in which the key concepts are located. The test questions generated can be multiple-choice, yes/no, fill-in-the-blanks, short-answer, or essay-type questions. If students provide written answers, the system can automatically compare their answers with the reference answer and provide a score (short-answer grading) by using deep learning technology and feedback. A smart evaluation system uses AI's deep learning technology, which offers accuracy that is closer to manual evaluation than that obtained through traditional machine learning methods. In summary, learning strategies and the design of learning activities and evaluation methods directly influence learning behavior and effectiveness. Therefore, they are critical key factors that strengthen learning analysis and improve both teaching effectiveness and students' learning effectiveness.

3. Contribution of papers to this special issue

Thirteen papers have been included in this special issue that can be classified into five categories. The papers include systematic overviews and literature reviews concerning precision education, framework and practice of precision education, diagnosis of behavior patterns through the application of precision education, identification of key concepts and quality ideas considering precision education, and prediction and factor analysis regarding precision education.

Three papers address systematic overviews and literature reviews concerning precision education, paper authored by Dirk Tempelaar, Bart Rienties and Quan Nguyen, entitled "The Contribution of Dispositional Learning Analytics to Precision Education"; paper authored by Xieling Chen, Di Zou, Haoran Xie, and Gary Cheng, entitled "Twenty years of personalized language learning: topic modelling and knowledge mapping"; and paper authored by Hui Luan and Chin-Chung Tsai, entitled "A review of using machine learning approaches for precision education"

Three papers address framework and practice of precision education, paper authored by Fuzheng Zhao, Gwo-Jen Hwang, and Chengjiu Yin, entitled "A Result Confirmation-based Learning Behavior Analysis Framework For Exploring The Hidden Reasons Behind Patterns and Strategies"; paper authored by Jiun-Yu Wu, Christopher C.Y. Yang, Chen-Hsuan Liao, and Mei-Wen Nian, entitled "Learning Analytics 2.0 for Precision Education: An Integrative Theoretical Framework of the Human and Machine Symbiotic Learning"; and paper authored by Tzu-Chi Yang, Yih-Lan Liu, and Li-Chun Wang, entitled "Using an institutional research perspective to predict undergraduate students' career decisions in the practice of precision education."

Three papers address the diagnosis of behavior patterns through the application of precision education, paper authored by Christopher C.Y. Yang, Irene Y.L. Chen and Hiroaki Ogata, entitled "Toward Precision Education: Educational Data Mining and Learning Analytics for Identifying Students' Learning Patterns with Ebook Systems"; paper authored by Xuanqi Feng and Masanori Yamada, entitled "An Analytical Approach for Detecting and Explaining the Learning Path Patterns of an Informal Learning Game"; and paper authored by Mehmet Kokoç, Gökhan Akçapinar, and Mohammad Nehal Hasnine, entitled "Unfolding Students' Online Assignment Submission Behavioral Patterns using Temporal Learning Analytics."

Two papers address the identification of key concepts and quality ideas considering precision education, paper authored by Albert C.M. Yang, Irene Y.L. Chen, Brendan Flanagan, and Hiroaki Ogata, entitled "From Human-grading to Machine-grading: Automatic Diagnosis of e-Book Highlighting in Precision Education", and paper authored by Alwyn Vwen Yen Lee, entitled "Determining Quality and Distribution of Ideas in Online Classroom Talk using Learning Analytics and Machine Learning."

Two papers address the prediction and factor analysis regarding precision education, paper authored by Hui-Chen Lin, Yun-Fang Tu, Gwo-Jen Hwang and Hsin Huang, entitled "From Precision Education to Precision Medicine: Factors Affecting Medical Staff's Intention to Learn to Use AI Applications in Hospitals", and paper authored by Feifei Han and Robert A. Ellis, entitled "Predicting Students' Academic Performance by Their Online Learning Patterns in a Blended Course: To What Extent Is a Theory-driven Approach and a Data-driven Approach Consistent?"

4. Conclusion and future research: Human-centered AI in education

As addressed in the introduction, taking humans into consideration is a challenge for AI; therefore, the application of AI in education should be human centered. Human-centered AI can be interpreted from two perspectives: AI under the control of humans (Shneiderman, 2020) and AI concerning the human condition (Stanford HAI, 2020). Considering the perspective of human-controlled AI, AI can be distinguished according to the degree of human control involved. One level concerns complete control by humans, with AI only assisting automation, and the other level concerns human autonomy that is completely determined by AI. Human-controlled AI leverages the collaboration between human control and AI automation to boost human productivity with offering a high level of reliability, safety, and trust (Shneiderman, 2020). From the point of view of AI concerning the human condition, AI algorithms have to take humanity as the main consideration, perform explainable and interpretable computation and judgement processes, and continuously adjust AI algorithms by considering the human context and societal phenomena to augment human intelligence with machine intelligence, thereby contributing to the welfare of humankind.

Research on human-centered AI advocates an in-depth dialogue between researchers from diversity of thought, genders, ethnicity, cultures, and disciplines to facilitate better understanding of human-centered AI. Beneficial interactions between researchers can promote the adoption of human-centered AI in education by augmenting human intelligence with machine intelligence. For human-centered AI, research topics concern governance and policy, including biases in AI algorithms, use and misuse of AI, the societal impact of AI, AI in governance, AI governance, AI risk management, AI accountability, and AI self-surveillance. The research topics related to technology and practice include explainable AI, interpretable machine learning, flexibility and contextual understanding by humans, explanation and comprehension by humans, intelligent agents (assistants), automated conversational robots (Chabot), AI-enabled personalization, intelligent tutoring systems, student and teacher modeling for smart learning and teaching, smart content, learning pathway and recommendations, differentiated and individualized learning, intelligent assessment and evaluations, automated question generation, automated grading, and plagiarism detection.

AI has the potential to educate, train, and increase human performance, in turn making humans better at their tasks and activities. Proper application of AI can enable human welfare through various means, such as by improving the productivity of food, health, water, education, and energy services. However, misuse of AI because of algorithm bias and lack of governance could inhibit human rights and result in inequality pertaining to job opportunities, gender, and race (Vinuesa et al., 2020). AI may be able to imitate human emotions; however, imitating human feelings is difficult for AI. Human emotions are triggered by hormonal changes produced because of physiological and chemical changes, and feelings are an internal cognitive sensation. The ability to self-reflect and judge in face of psychological conflicts and contradictions, that is, inner cognition, drives humans to execute a series of reflective cycles through learning, unlearn, and relearn. Human beings are not perfect, and only inner beauty holds substance. Human beings should accept that their present selves are the best version of themselves and adjust emotionally to see the invisible mind through the visible world. Such a perception will allow humans to engage with the beauty of the mind.

Acknowledgement

The guest-editor would like to express sincere thankfulness to Prof. Nian-Shing Chen of National Yunlin University of Science and Technology, Taiwan, and Prof. Maiga Chang of Athabasca University, Canada, for their inspiring advices to this special issue. Given the high volume of submissions, this special issue runs with an acceptance rate lower than 20%, only a few papers that most related to “Precision education” are included. We truly appreciate all the authors and reviewers for their valuable contribution and support towards the success of this Special Issue and ET&S.

References

Lu, O., Huang, A., Huang, J., Lin, A., Ogata, H., & Yang, S. J. H. (2018). Applying learning analytics for the early prediction of students’ academic performance in blended learning. *Educational Technology & Society, 21*(2), 220-232.

Luan, H., Geczy, P., Lai, H., Gobert, J., Yang, S. J. H., Ogata, H., Baltes, J., Guerra, R., Li, P., & Tsai, C.C. (2020). Challenges and future directions of Big data and artificial intelligence in education. *Frontiers in Psychology, 11*. doi:10.3389/fpsyg.2020.580820

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction, 36*(6), 495-504.

Stanford HAI (2020). *Stanford Institute for Human-Centered Artificial Intelligence*. Retrieved from https://hai.stanford.edu/

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmar, M., & Nerini, F. F. (2020). The Role of artificial intelligence in achieving the sustainable development goals. *Nature Communication, 11*(233), 1-10. doi:10.1038/s41467-019-14108-y

Yang, S. J. H. (2019). Precision education: New challenges for AI in education [conference keynote]. In *Proceedings of the 27th International Conference on Computers in Education (ICCE)* (pp. XXVII-XXVIII). Kenting, Taiwan: Asia-Pacific Society for Computers in Education (APSCE).

The Contribution of Dispositional Learning Analytics to Precision Education

Dirk Tempelaar[1*], Bart Rienties[2] and Quan Nguyen[3]
[1]Maastricht University, School of Business and Economics, The Netherlands // [2]The Open University, Institute of Educational Technology, UK // [3]School of Information, University of Michigan, Ann Arbor, MI, USA //
d.tempelaar@maastrichtuniversity.nl // bart.rienties@open.ac.uk // quanngu@umich.edu
[*]Corresponding author

ABSTRACT: Precision education requires two equally important conditions: accurate predictions of academic performance based on early observations of the learning process and the availability of relevant educational intervention options. The field of learning analytics (LA) has made important contributions to the realisation of the first condition, especially in the context of blended learning and online learning. Prediction models that use data from institutional information systems and logs of learning management systems have gained a good reputation in predicting underperformance and dropout risk. However, less progress is made in resolving the second condition: applying LA generated feedback to design educational interventions. In our contribution, we make a plea for applying dispositional learning analytics (DLA) to make LA precise and actionable. DLA combines learning data, as in LA, with learners' disposition data measured through self-report surveys. The advantage of DLA is twofold: first, it improves the accuracy of prediction, specifically early in the module, when limited LMS trace data are available. Second, the main benefit of DLA is in the design of effective interventions: interventions that focus on addressing individual learning dispositions that are less developed but important for being successful in the module. We provide an empirical analysis of DLA in an introductory mathematics module, demonstrating the important role that a broad range of learning dispositions can play in realising precision education.

Keywords: Blended learning, Dispositional learning analytics, Educational intervention, Flipped learning, Precision education

1. Introduction

The "engineers" approach to education' seems to be a good way to define precision education casually. A more formal definition, derived from Yang (2019), states that precision education aims to identify at-risk students as early as possible and provide timely intervention through diagnosis, prediction, treatment, and prevention. These four components imply that the diagnosis of students' learning patterns, the prediction of students' learning outcomes, the treatment with learning strategy and activities, and prevention shape the four main research topics in precision education. The metaphor of engineering science does bring some more insights, whereby both system theory and control theory are important building blocks. System theory is engaged with the description of systems by mathematical models that enable the estimation of the state of the system and the prediction of the outcome of the system resulting from that state (Padulo & Arbib, 1974). Control theory is based on system theory and adds the dimension of steering the state of the system towards a predefined, desired constellation (Padulo & Arbib, 1974).

When comparing educational theory with system theory, precision education stands for the added dimension of being in control with regard to the educational attainments. To be "in control" in the system theory context, two so-called structural properties are required: observability and controllability (Padulo & Arbib, 1974). Controllability, also called reachability, refers to the interaction of system inputs and the system state. A system is controllable if, with proper input, every possible state can be reached. Observability relates to the interaction of state and output. A system is observable if the measurements of both inputs and outputs contain sufficient information to reconstruct any state (reconstructability is the alternative term for observability). Looking at the definition of precision education provided by Yang (2019), one can identify both aspects. Diagnosis and prediction refer to the observability aspect, whereas treatment and prevention refer to the controllability aspect.

Although precision education is a young discipline, different interpretations can be found as to what its focus is. Often, that difference is about a focus on both aspects of controllability and observability versus a focus on observability only. Next, most empirical research into precision education concentrates on the ability to predict students' academic performance and to identify students at risk for dropout. For example, Hart (2016, p. 209) suggested a focus on "creating data specifically for gaining a better understanding of the classification of learning disabilities at the individual level." Lian and Sangarun (2017), in the context of language education, defined the "precision project" in terms of providing accurate, detailed, timely, adaptive, and contextualised personalised data to facilitate intervention. Three empirical studies, Lu et al. (2018), Hsiao et al. (2019), and Huang et al. (2020), although acknowledging the role of intervention in precision education by phrasing that "*learning analytics is a conceptual framework and as a part of our Precision education used to analyze and predict students' performance and provide timely interventions based on student learning profiles*" (Lu et al., 2018, p. 221), focus on the role of learning analytics in providing accurate predictions of academic performance.

We make a plea for equal focus on prediction and intervention, observability and controllability in this contribution. We will materialise that plea by adopting a dispositional learning analytics (DLA) framework. It is our conjecture that adding dispositions of learners to learning analytics is a small step to take that may bring considerable benefits in terms of increasing observability, creating a better picture of the features of a learning issue, as well as increasing controllability, transforming predictions of poor performance or risk of dropout into meaningful interventions.

2. Learning analytics and dispositional learning analytics

Learning analytics (LA) refers to "the analysis and interpretation of educational data, such as the logs recorded in learning management systems, the interactive contents recorded in online discussion forums, or the learning process captured on video, to provide constructive feedback to learners, instructors or educational policymakers" (Hwang et al., 2018, p. 134). At the early stage of learning analytics, many scholars have focused on building predictive models based on data extracted from both institutional student information systems (SIS) and digital platforms that organise and facilitate learning, such as learning management systems and e-tutorials (LMS, taken them together). While these studies provide important markers on the potential of LA in education (Viberg et al., 2018), the findings were rather limited to the descriptive function of LA, which is mostly based on the "measurement, collection, analysis and reporting of data about learners and their contexts" (Siemens & Gašević, 2012, p. 1). Given the rigidity of SIS and LMS data, educators may encounter difficulties in designing pedagogically informed interventions. To overcome this shortcoming, Buckingham Shum and Crick (2012) proposed the Dispositional LA (DLA) infrastructure that combines learning data (i.e., generated in learning activities through traces of an LMS) with learner data (e.g., student dispositions, values, and attitudes measured through self-report surveys). Typical stakeholders of DLA applications are both students and teachers/tutors: these applications can be positioned at the meso- and micro-level (Ifenthaler, 2015).

It is our conjecture that, especially in precision education, these dispositions of learners are a crucial add-on to common LA functions. They are the dispositions that make the feedback both precise and actionable (Gašević et al., 2015; Tempelaar et al., 2017a). The term "actionable feedback," introduced into the LA community by Gašević et al. (2015), is crucial in pinpointing the aspect of precision education, not by default present in LA applications. Although the concept is well accepted in the LA community, not much empirical research has been published focusing on the extent to which LA feedback is actionable. Some exceptions are, for instance, Gašević et al. (2017), in which students' learning strategies are derived from log data, or Abdous et al. (2012), where learning behaviours in terms of frequency of chat messages and platform questions are analysed. Indeed, in a large scale application of LA of 1159 teachers in 231 courses containing 23K students over four years, Herodotou et al. (2020) found that although teachers appreciate rich predictive learning analytics data of whom might be at risk of dropping out, several groups of teachers struggled to put this advice into action.

Adding dispositions opens the perspective of enriching the LA learning feedback. Rather than intervening with the simple message to Maria "please catch up, you are lagging behind," the intervention can now contain some actionable aspects as "We see that your performance is not as good as most of your peers and, at the same time, we observe that you view the videos only after the class session has taken place, rather than before. It might be better to try viewing these videos before meeting in class." Still, this feedback is unidimensional in addressing one singular aspect of learning behaviour, and if no deviant behaviour can be detected in the learning strategies of the student lagging behind, we are still empty-handed with regard to actionable feedback. Therefore, the aim of DLA is to introduce a multidimensional perspective of learning dispositions into the practice of LA by identifying which constellations of personality characteristics go together with the risk of dropout or the prediction of underperformance. In Buckingham Shum and Deakin Crick (2012) and Buckingham Shum and Ferguson (2012) the source of learner data is found in the use of a dedicated survey instrument specifically developed to identify learning power (Deakin Crick et al., 2015): the mix of dispositions, experiences, social relations, values, and attitudes that influence the engagement with learning.

Sharing a similar systems' approach as described in Deakin Crick et al. (2015), we propose to operationalize dispositions with the help of instruments developed in the context of contemporary educational research, as to make the connection with educational theory and pedagogical interventions as strong as possible. Somewhat preluding on the outcomes of the empirical part of our contribution, the following illustrates the role of a broad spectrum of learning dispositions, which may make learning feedback actionable. Continuing with the example described above, when a student lags, a traditional LA application cannot go any further than providing a warning: "please catch up!"

If we extend the LA application as in Gašević et al. (2017) by deriving learning strategies, we can extend the feedback with reference to suboptimal strategies applied. However, this feedback is also at the symptom level and may trigger symptom management rather than looking for potential causes of these symptoms. In other words, the learning feedback is not that actionable as to suggest concrete forms of intervention, and the controllability condition is not yet fulfilled. In previous research of the authors (Nguyen et al., 2016; Rienties et al., 2019; Tempelaar et al., 2015; Tempelaar et al., 2017a, Tempelaar et al., 2017b; Tempelaar et al., 2018a, Tempelaar et al., 2018b; Tempelaar et al., 2019) and again in the empirical part of this contribution, we found that different constellations of learning dispositions might be associated with underperformance and lagging behind in the learning process. One such constellation may be based on personality characteristics of motivational types, such as disengagement. A very different constellation can be based on learning regulation strategies: the balance between self-regulation and external regulation of learning.

While typically education strives to achieve high levels of self-regulated learning, our empirical studies indicated that a minimum of external regulation, open up to the teachers' advice and the structure of the curriculum to a certain level, is indispensable. Both types of students, the disengaged and the too strong self-regulated, demonstrate the same symptoms: not being very active in doing the scheduled learning activities, lagging in what is measured (Tempelaar et al., 2015). However, the independent learner who makes his own way (indeed, more often his than her) is certainly not disengaged and requires feedback and intervention of a very different type than the disengaged learner. It is here that the dispositions component of DLA can be the crux to precision education, as we intend to showcase in the empirical part of our contribution. Rather than formulating specific research hypotheses, we will design our empirical study around this open research question: how can DLA contribute to prediction and intervention, to both dimensions of observability and controllability? Given the strong focus of most LA applications on observability issues, on the derivation of prediction models, we will emphasise the component of controllability in our study: in what respect can DLA better than LA facilitate the design of learning interventions directed at taking observed learning obstacles.

3. The educational context of the empirical study: blended learning and flipped classes

The learning context investigated in previous research by the authors as well as in the empirical part of this contribution is best described as large-scale introductory mathematics and statistics module using "blended" or "hybrid" learning in a business and economics university program in the Netherlands. Blended learning "combines online digital resources with traditional classroom activities and enables students to attain higher learning performance through well-defined interactive strategies involving online and traditional learning activities" (Lu et al., 2018, p. 220). The main learning component in our blend is face-to-face: Problem-Based Learning (PBL), in small groups (14 students), coached by a content expert tutor (Non & Tempelaar, 2016; Williams et al., 2016). Participation in these tutorial group sessions is required. Optional is the online component of the blend: the use of two e-tutorials or online learning and practising environments: the e-tutorial SOWISO to learn and practice mathematics (https://sowiso.nl/en/) and the e-tutorial MyStatLab to learn and practice statistics (https://www.pearsonmylabandmastering.com/northamerica/mystatlab/) (see Tempelaar et al., 2015; Tempelaar et al., 2017a, Tempelaar et al., 2017b; Tempelaar et al., 2018a, Tempelaar et al., 2018b; Tempelaar et al., 2019). This choice is based on the philosophy of student-centred education, placing the responsibility for making educational choices primarily on the student. Since most of the learning takes place in self-study outside class using the e-tutorials or other learning materials and class time is used to discuss solving advanced problems, the instructional format is best characterised as a flipped-class design (Hsiao et al., 2019; Lin & Hwang, 2018; Williams et al., 2016). The use of e-tutorials and achieving good scores in the practising modes of both e-tutorials is stimulated by making bonus points available for good performance in the quizzes. Quizzes are taken every two weeks and consist of items that are drawn from the same item pools applied in the practising mode. We chose this particular constellation as it stimulates students with limited prior knowledge to make intensive use of the e-tutorials. The bonus is maximised to 20% of what one can score in the exam.

The student-centred nature of the instructional design requires, first and foremost, adequate actionable feedback to students so that they can monitor their study progress and topic mastery. The provision of relevant feedback starts on the first day of the module when students take two diagnostic entry tests for mathematics and statistics. Feedback from these entry tests provides a first signal of the importance of using the e-tutorials. Next, the e-tutorials SOWISO and MyStatLab take over the monitoring function: at any time, students can see their performance in the practice sessions, their progress in preparing for the next quiz, and detailed feedback on their completed quizzes, all in the absolute and relative (to their peers) sense.

Our program is characterised by a large diversity in the student population: only about 20% of the students are educated in the Dutch high school system, most students are international, with a large share of European nationalities: no more than 5% of students are from outside Europe. High school systems in Europe differ strongly, most notably in the teaching of mathematics and statistics. Therefore, it is crucial that the first module is flexible and allows individual learning paths (Non & Tempelaar, 2016; Williams et al., 2016).

Learning dispositions measured at the start of the course were of affective, behavioural, and cognitive types (Rienties et al., 2019). The surveys had a prime role in supplying students with an individual data set required for doing a statistical project, resulting in a full response.

4. Methods for the current study

4.1. Participants

In the empirical part of this contribution, we investigated DLA's potentials applied to two cohorts of first-year students in our program: the cohorts of academic years 18/19 and 19/20. Profiling of students took place on the basis of a quartile split of the scores of all students who participated in the first quiz: 2,261 students (since the collection of survey data is part of a mandatory assignment, full data are available for all these students). Of these students, 40% were female, 60% male (with *Female* as indicator variable), and 35% followed the advanced mathematics track in high school education (with *MathMajor* as indicator variable).

4.2. E-tutorial trace data

Although students learned in two different e-tutorial systems, SOWISO for mathematics and MyStatLab for statistics, we focus the analysis on data gathered from the SOWISO platform since that platform allows us to trace every individual learning activity as a time-stamped record in the database of loggings. Students spend an average of 27 hours in SOWISO or 3.5 hours per week, out of the ten hours per week available to learn mathematics (that includes both self-study and class time). Learning can be divided into three consecutive learning phases. The first learning phase prepares the weekly tutorial session: students are expected to enter these sessions well prepared by self-studying the weeks' topic in advance so that the session itself can be used to solve advanced problems. The second learning phase starts after the tutorial session took place and runs until the bi-weekly quiz session. In phase two, students prepare the quizzes that bring them a bonus score. The third and last learning phase is the examinations' preparation: it starts after writing the quiz and continues until the writing of the exam begins. In each learning phase and for each of the seven weekly topics covered in the program, the students' mastery in the practice mode of the e-tutorial is measured: the proportion of problems successfully solved. Three learning phases and seven weekly topics imply a total of 20 mastery measures: *TGMasteryTopicWk1*, *QzMasteryTopicWk1*, *ExMasteryTopicWk1*, to *ExMasteryTopicWk7*, where *TG*, *Qz*, and *Ex* refer to tutorial group session (first learning phase), quiz session (second learning phase), and examination (third learning phase) (the last weekly topic is not included in any quiz).

4.3. Performance data

Following the focus on mathematics learning, as explained above, two different types of performance indicators are available: exam score and quiz scores. There are three bi-weekly quizzes, covering the topics of the first two weeks (*MathQuiz1*), the second two weeks (*MathQuiz2*), and the third two weeks (*MathQuiz3*) (the last week topic not covered in any quiz). *MathExam* represents the score for the mathematics component in the final examination. Since this study aims to investigate the potential of early prediction of performance, most of our analysis is focused on the role of the first quiz that is administered in the third week of the module in predicting performance, and the relationships between learning dispositions and the score achieved in that first quiz:

MathQuiz1. The module starts with a diagnostic entry test, producing *EntryTest* as a measure of prior knowledge.

4.4. Disposition data

Five different instruments were applied to operationalize learning dispositions, all documented in full detail in previous studies (Tempelaar et al., 2015, Tempelaar et al., 2017a, Tempelaar et al., 2017b; Tempelaar et al., 2018a, Tempelaar et al., 2018b; Tempelaar et al., 2019). For space limitations, we limit the current description to the identification of survey scales adopted and refer to the above sources for a full elaboration.

Individual approaches to cognitive learning processing strategies and metacognitive learning regulation strategies were based on Vermunt's (1996) learning styles instrument. Processing strategies can be ordered from surface to deep learning approaches: *Memorising* and rehearsing and *Analysing* are two scales that represent different aspects of surface learning. *Relating* and structuring and *Critical* processing are two scales that represent different aspects of deep learning. *Concrete* processing is separate from this continuum and represents the tendency to learn strategically. Regulation strategies are decomposed into self and external regulation: Self-regulation of learning processes and results (*SelfRegProc*), Self-regulation of learning content (*SelfRegCont*), External regulation of learning processes (*ExtRegProc*), and External regulation of learning results (*ExtRegRes*), with Lack of regulation (*LackReg*) indicating a lack of regulation of any type.

Attitudes and beliefs toward learning quantitative topics were assessed with the SATS instrument (Tempelaar et al., 2007). It distinguishes Affect, cognitive competence (*CognComp*), *Value*, expected difficulty in learning, reversed (*NoDifficulty*), *Interest*, and planned *Effort*.

Learning emotions, both epistemic and activity type were measured on the basis of instruments developed by Pekrun (Pekrun & Linnenbrink-Garcia, 2012). Epistemic emotions are composed of positive emotions, *Curious* and *Excited*, negative emotions *Confused*, *Anxious*, *Frustrated*, and *Bored*, and the neutral emotion *Surprised* (Pekrun et al., 2017). Activity emotions were measured the module halfway, in the fourth week. Although these emotions are strongly associated with performance measures, we left them out of the analysis, given the focus of early prediction.

The instrument Motivation and Engagement Wheel (Martin, 2007) breaks down learning cognitions and learning behaviours into four categories of adaptive versus maladaptive types and cognitive versus behavioural types. *Self-belief*, value of school (*ValueSchool*), and learning focus (*LearnFocus*) shape the adaptive, cognitive factors, or cognitive boosters. Planning, task management (*TaskManagm*), and *Persistence* shape the behavioural boosters. Mufflers, the maladaptive, cognitive factors are *Anxiety*, failure avoidance (*FailureAvoid*), and uncertain control (*UncertainControl*), while self-sabotage (*SelfSabotage*) and *Disengagement* are the maladaptive, behavioural factors or guzzlers.

A recently developed 4x2 achievement goal framework by Elliott and coauthors (Elliot et al., 2015) was applied to include the self-perceived goal-setting behaviour of students. The instrument distinguishes two valence dimensions: approach and avoid, and four goal definition dimensions: task-based competence (striving to do the task correctly), self-based competence (do better than before), other-based competence (do better than others) and potential-based (to do the best one can) competence, resulting in eight scales: *TAP* or Task-Approach, *TAV* or Task-Avoid, *SAP* or Self-Approach, *SAV* or Self-Avoid, *OAP* or Other-Approach, *OAV* or Other-Avoid, *PAP* or Potential-Approach, and *PAV* or Potential-Avoid achievement goals.

4.5. Analyses

In this study, we choose to profile students in a very simple way: by quartile split of the first quiz score: *MathQuiz1*. In previous research (Tempelaar et al., 2015, Tempelaar et al., 2017a, Tempelaar et al., 2017b; Tempelaar et al., 2018a, Tempelaar et al., 2018b; Tempelaar et al., 2019), more advanced statistical techniques as cluster analysis or latent class analysis served the role of distinguishing students' different learning profiles. However, to distinguish student profiles with different dispositional characteristics, we do not need advanced statistical methods, which will be illustrated by our simple quartile split. In this study, the prime aim of predictive modelling is directed at predicting academic performance and distinguishing different groups or profiles that are best helped with different types of learning feedback based on different dispositional profiles (such as raising confidence for students being failure avoidant or helping to organise the study in students who lack planning skills). Since the first quiz score is the dominant early predictor of final course performance (see next section), a simple quartile split into four different profiles is adequate. Profile means are compared with ANOVA analyses. Differences in means are classified as statistically significant when p-values are below .01.

5. Experimental results and discussion

5.1. Prediction equations of final performance

The first step in the analysis is to consider the predictive power of alternative sets of predictors to explain the final exams' performance. Given that our investigations are directed at finding prediction models that allow for timely interventions, we restricted predictors to constructs measured before the start of the module or in the first three weeks of the module (allowing ample time for intervention in the later five weeks of the module). The alternative sets of predictor variables are:

- Demographics measured before the start of the module: gender, prior education, score in diagnostic entry test and dummy for year.
- Demographics plus the learning dispositions discussed above in the subsection of disposition data, all measured before the start or at the start of the module.
- Demographics plus learning dispositions plus trace variables collected in the first three weeks of the module: learning mastery in the e-tutorial for each of the topics covered in the first three weeks of the program, plus the score in the first quiz taking place in the third week of the module.
- Space prevents us to report the full regression models (mainly due to the 34 dispositional variables in the prediction model). Therefore, Table 1 restricts reporting the predictive power of the several models.

Table 1. Predictive power of three prediction models of final performance

Predictor set	R^2
demographics	0.184
demographics + dispositions	0.305
demographics + dispositions + traces	0.411

Table 1 emphasises one of the roles of dispositional variables (beyond their role in the actionable feedback): they empower the prediction equations raising predictive power of diagnostic entry test and prior education (gender is insignificant as predictor) with 12%. Trace variables add another 10.5%. Within the set of trace variables, the score in the first Quiz is the dominant predictor. In itself, MathQuiz1 explains 28.4% of variation in MathExam, leaving no more than a modest 12.7% of additional explained variation for all demographic, dispositional and trace variables. This outcome is fully in line with findings in previous studies (Tempelaar et al., 2015, Tempelaar et al., 2017a, Tempelaar et al., 2017b; Tempelaar et al., 2018a, Tempelaar et al., 2018b; Tempelaar et al., 2019): it is only in the early start of the module that demographic variables, dispositional variables or learning activity variables measured by traces in learning environments contribute substantially to the prediction of learning performance. Very early learning interventions, taking place before the first measures of cognitive types, such as quiz scores, are collected, are best based on the full set of predictors. However, as soon as these cognitive measures are included, they dominate all other types of predictors in the prediction equations.

5.2. Student profiles by quartile split

Four student profiles are determined as the outcome of a quartile split of *MathQuiz1* score. That score expressed as a proportion, the three quartiles are .458, .667 and .804. The four quarters have unequal size, due to the stepwise nature of the *MathQuiz1* scores. Q1 counts 473 students scoring below 0.458; Q2 counts 648 students scoring between .458 and .667; Q3 counts 573 students scoring between .667 and .804, and Q4 counts 567 students scoring .804 or beyond. Average scores of all four quarters for the three quiz scores and the final exam score, all expressed as proportions are depicted in Figure 1.

All profile differences are strongly statistically significant and substantial in size (ANOVA F-values > 200, p-values < 10^{-10}). Since the profiling is based on *MathQuiz1* scores, differences are largest for that variable, but also substantial in the other scores. Given that the passing norm is 55%, profile differences in passing rates are even higher. With most students in Q1, the first quarter, failing, and nearly all students in the third and fourth quarters, Q3 and Q4, passing.

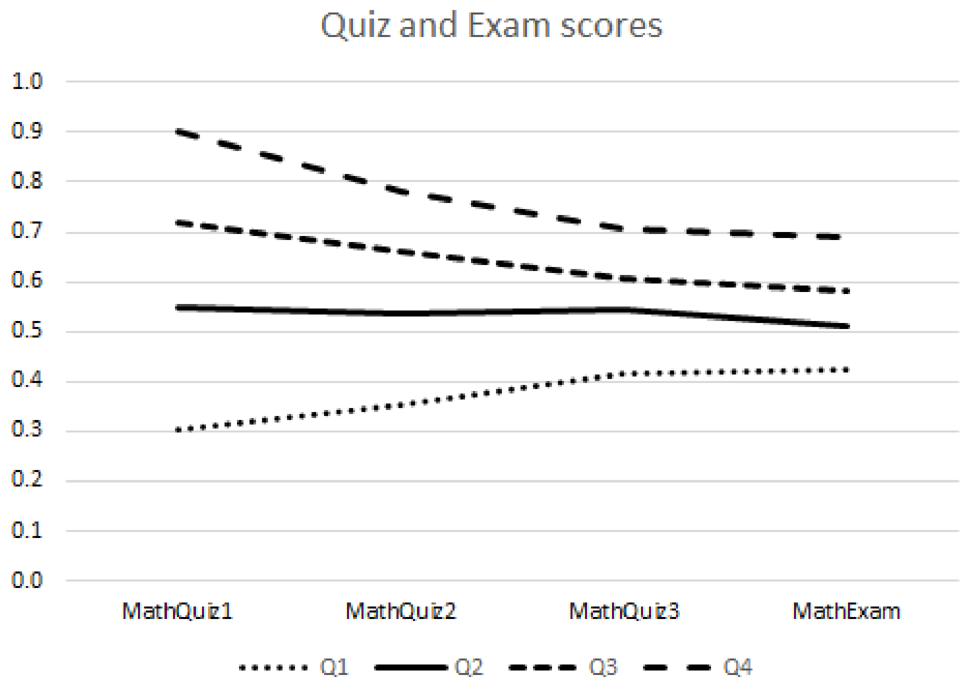


Figure 1. Profiling students by quartile split: Quiz and exam performance

5.3. Student profiles by quartile split: Traces of learning activity

In a traditional LA application, the first thing to check is if the profiling is validated by differences in levels of learning activity. Our application focuses on the mastery level defined as the proportion of successfully solved practice problems achieved in the e-tutorial in the first three weeks of the module. Every week has its own topic to be covered, so mastery levels refer to three different topics: see the three panels of Figure 2. In every topic, we distinguish three subsequent learning phases: mastery achieved in the preparation of the tutorial session (*TGMastery*), master achieved in the preparation of the quiz session (*QzMastery*), and mastery achieved in the preparation of the examination (*ExMastery*). Mastery levels are cumulative: mastery in subsequent phases add to the mastery achieved in previous ones.

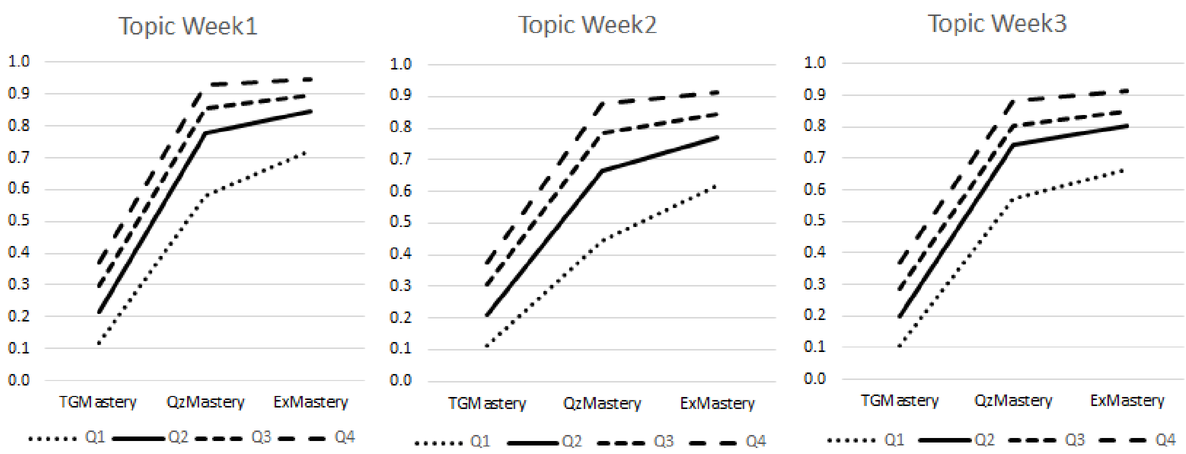


Figure 2. Profiling students by quartile split: Mastery by learning activity in the first three weeks

We observe considerable differences between profiles, all strongly statistically significant (ANOVA F-values > 50, p-values < 10^{-10}). Figure 2 suggests that most of the learning takes place in the second phase: after the tutorial session took place, in preparation for the quiz.

5.4. Student profiles by quartile split: Demographics

Prior education plays a major role in the explanation of the composition of the four different profiles. The most considerable differences are in the *MathMajor* variable: 58% of students in Q4 are educated in the science preparing track in high school, against no more than 15% in Q1 (ANOVA F = 81.5, p-value < 10^{-10}). These differences in prior education show up in the *EntryTest*: scores students achieve in the diagnostic entry test taken before the module starts (ANOVA F = 133.7, p-value < 10^{-10}). There are no gender differences between the four quarters: see Figure 3 (ANOVA F = 1.7, p-value = .162).

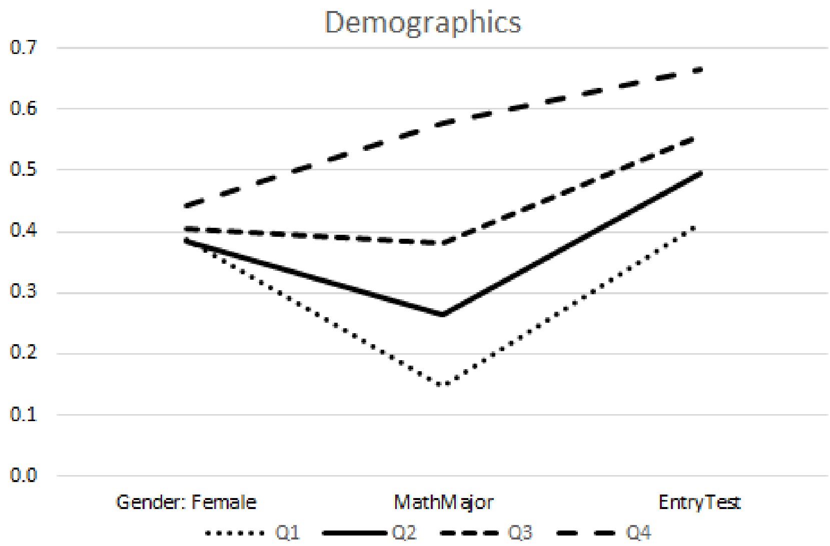


Figure 3. Profiling students by quartile split: Gender, prior education, and entry test score

5.5. Student profiles by quartile split: Learning processing and learning regulation strategies

Profile differences of learning processing and regulation strategies are at a much lower level, as Figure 4 illustrates. Statistical significant differences exist for *Relating* and structuring (ANOVA F = 4.0, p-value = .008), one of the deep learning components, for *ExtRegRes*, the external regulation of learning results (ANOVA F = 3.9, p-value = .009), and for Lack of regulation (ANOVA F = 10.6, p-value < 10^{-7}). Concerning learning processing: the higher the profile, the higher the level of deep learning, the lower the level of concrete learning. Concerning learning regulation: the higher the profile, the higher the level of external regulation of learning results, the lower the level of a lack of regulation.

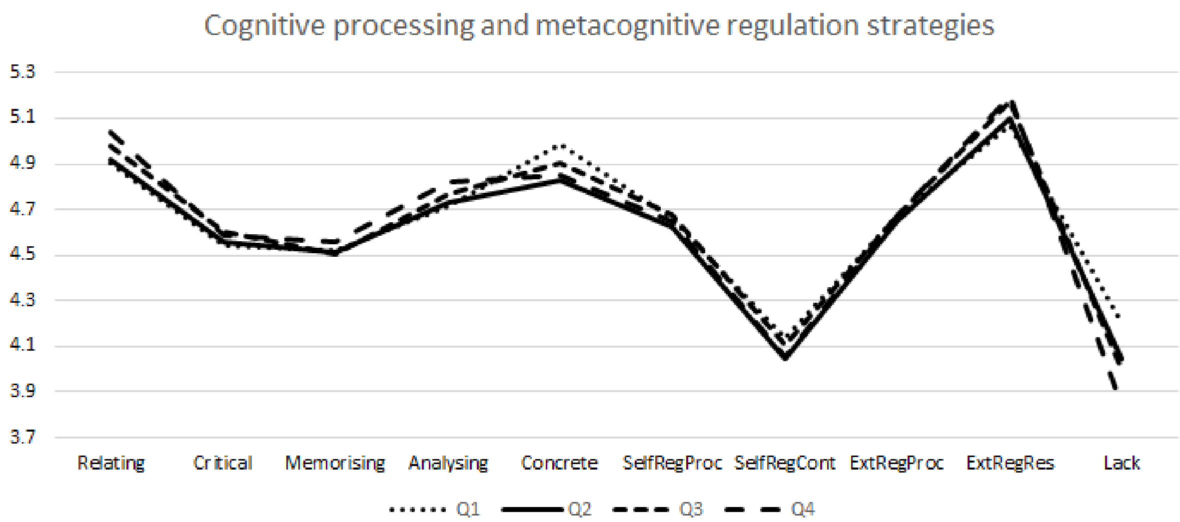


Figure 4. Profiling students by quartile split: learning processing and regulation

5.6. Student profiles by quartile split: Learning attitudes and beliefs

More considerable profile differences pop up when analysing attitudes and beliefs towards learning mathematics. All facets, except *Effort*, demonstrate statistically significant differences (ANOVA $F > 8.0$, p-values $< 10^{-4}$). However, the size of the differences is especially large for the *Affect* and *CognComp* variables, with levels in these attitudes increasing with higher quartiles. See Figure 5.

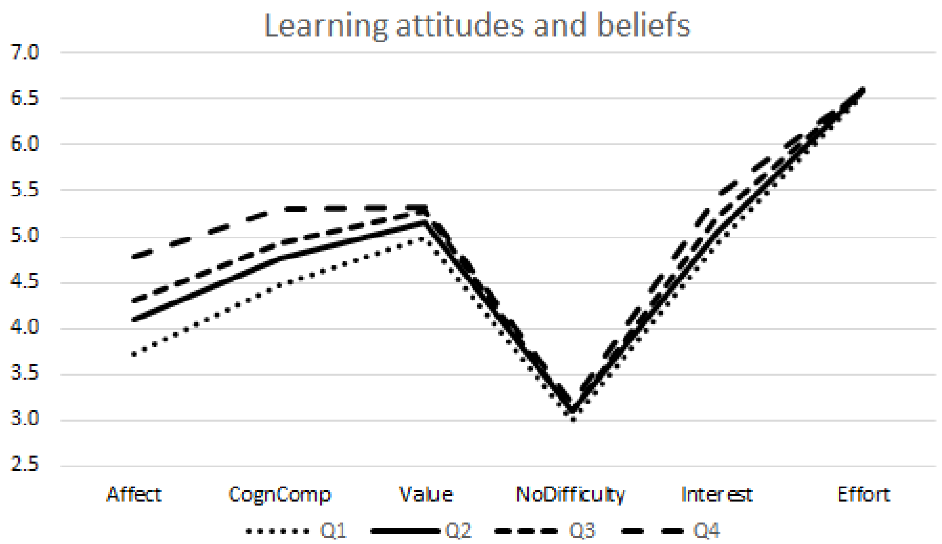


Figure 5. Profiling students by quartile split: Learning attitudes and beliefs

5.7. Student profiles by quartile split: Epistemic learning emotions

The valence dimension of epistemic emotion splits the graph of profile means into two mirrored patterns. Positive emotions *Curious* and *Excited* are positively related to the ordering of quarters; *Surprise*, hypothesised as a neutral emotion, is the only epistemic emotion without differences in profile means; all other differences are statistically significant (ANOVA $F > 10.4$, p-values $< 10^{-7}$). Negative epistemic emotions *Confused*, *Anxious*, *Frustrated*, and *Bored* demonstrate levels that are inversely related to the quarters' order with substantial profile differences: see Figure 6.

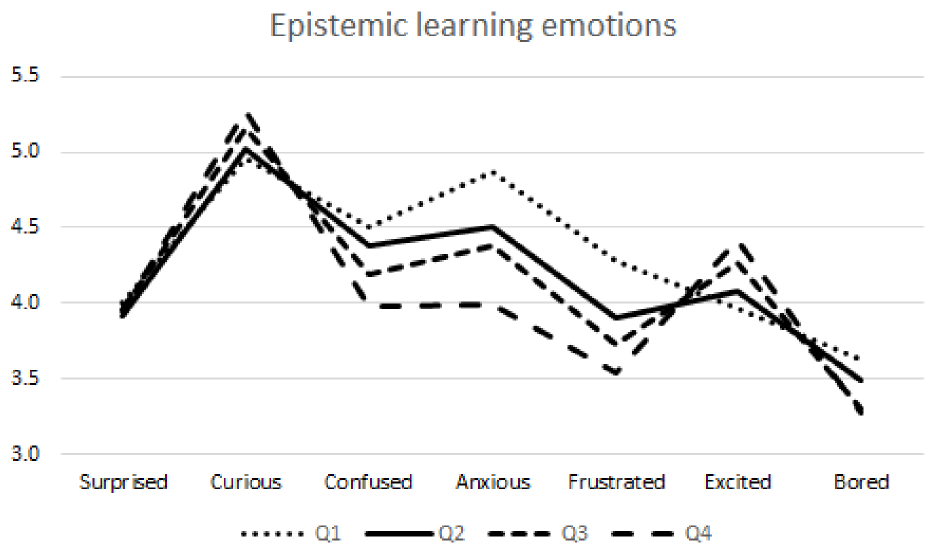


Figure 6. Profiling students by quartile split: Epistemic learning emotions

5.8. Student profiles by quartile split: adaptive motivation and engagement

Cognitive and behavioural motivation and engagement constructs of adaptive type demonstrate different patterns. The cognitive scale *Self-belief* demonstrates a small but significant profile difference (ANOVA $F = 5.8$, p-value = .0006); *LearnFocus* (ANOVA $F = 3.5$, p-value=.015) and *Value School* do not (ANOVA $F = .50$, p-values = .684). The behavioural constructs *Planning* and *Persistence* demonstrate profile differences of more substantial size, in the direction that higher ordered quarters reach higher mean levels (ANOVA $F = 6.0$ and 15.1, p-values = .0005 and 10^{-9}). No mean differences exist for *TaskManagm* (ANOVA $F = .35$, p-value = .791): see Figure 7.

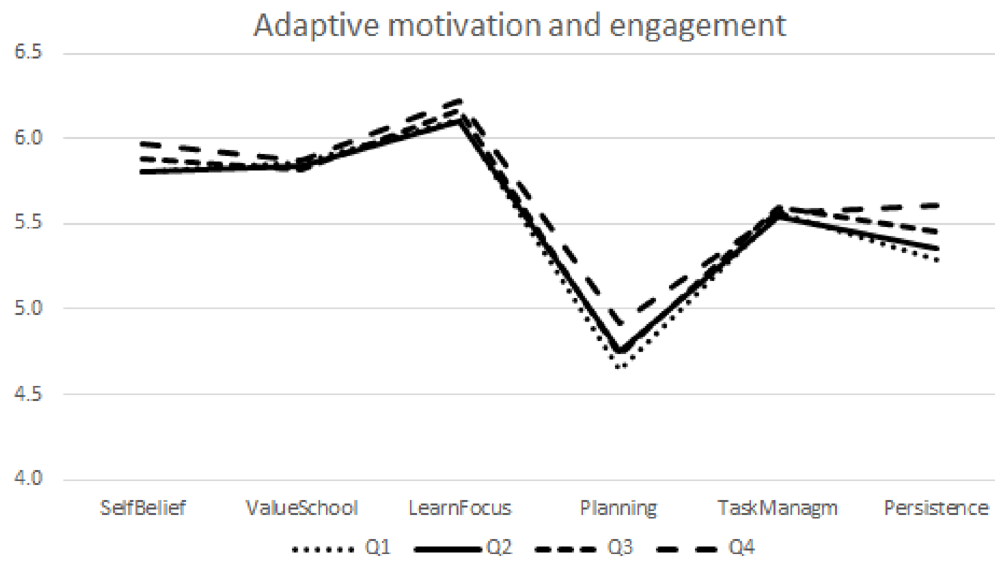


Figure 7. Profiling students by quartile split: Adaptive motivation and engagement

5.9. Student profiles by quartile split: maladaptive motivation and engagement

The same breakdown of cognitive and behavioural mean profile levels is visible in the next figure, Figure 8, providing patterns for maladaptive motivation and engagement constructs. The differences in profile means of the maladaptive cognitive constructs, *Anxiety* and *UncertainControl* (ANOVA $F = 10.6$ and 12.3, p-values < 10^{-6} and 10^{-7}), as well as those of the behavioural constructs *SelfSabotage* and *Disengagement* (ANOVA $F = 15.7$ and 4.2, p-values < 10^{-9} and .005), are all statistically significant but modest in size. Comparing Figure 7, we see that the pattern is inverted: the first quarter scores highest, the fourth quarter scores lowest.

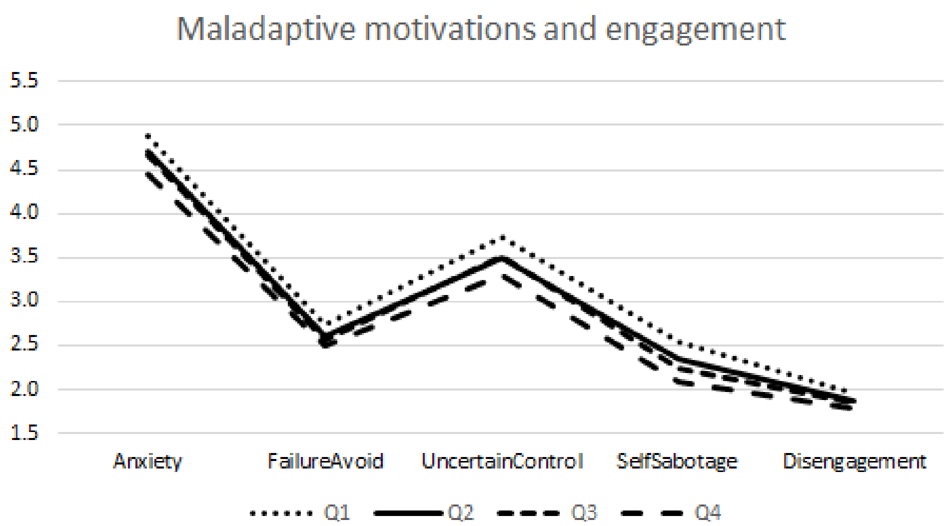


Figure 8. Profiling students by quartile split: Maladaptive motivation and engagement

5.10. Student profiles by quartile split: goal orientations

The fifth learning dispositions instrument administered is that of goal orientations. It provides a remarkable pattern of profile means: see Figure 9. First, profile differences between the four definitions of achievement goals, task-based, self-based, other-based, and potential-based, tend to exceed the differences observed in the valence dimension: approach versus avoidance. Next: differences in the self-related goals are absent (ANOVA F = .90 and .03, p-values = .443 and .992). The other three aspects of the definition, task-based goals (ANOVA F = 6.1 and 4.7, p-values < .001 and .003), other-based goals (ANOVA F = 10.4 and 5.6, p-values < 10^{-6} and .001), and potential-based goals (ANOVA F = 5.4 and 4.7, p-values = .001 and .003), demonstrate significant differences that are substantial in the case of other-based goals. That is: high-achieving students distinguish most from low-achieving students in their competitive learning motivation.

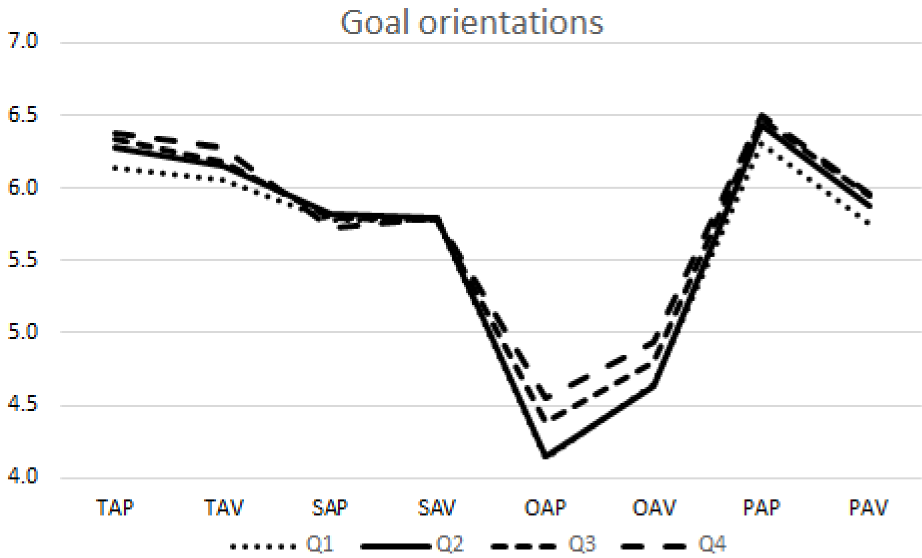


Figure 9. Profiling students by quartile split: Goal orientations

6. Conclusions

The "LA component" of our analysis results in outcomes that are representative for many empirical LA studies that focus on the prediction of academic performance: Lu et al. (2018), Hsiao et al. (2019), and Huang et al. (2020). Yes, there are several "early predictors" of module performance, of which in our study, the first quiz score dominates predictive power. Still, other variables like demographics and learning dispositions do contribute to explained variation. And yes again, trace measures of students' learning activity behaviours are associated with the first quiz score, suggesting an intervention directed at students with low levels in the e-tutorial in the first weeks of the module. Since we collected trace data at the individual level, such learning activity based feedback to students need not be at the profile level but can be tailored towards individual levels, as precision education suggests.

However, does such an intervention approach go beyond the symptom level? Are low activity levels really the cause of underperformance, or no more than a symptom of still hidden causes? In the "D" component of our DLA approach, we demonstrated that underperformance concerning our early performance measure is associated with a whole range of affective, behavioural, and cognitive learning dispositions. Learning attitudes, cognitive learning processing strategies, metacognitive learning regulation strategies, epistemic learning emotions, goal orientations, motivation and engagement, all have their role in explaining variation in the first quiz score. Since our module is the first module students take in the first term of the first year of university, and the dispositions are measured at the start of the module, these dispositions represent learning approaches and beliefs students acquired during six years of high school education and upon transferring to university, bring into our university classes. These suggest being one of the true causes of underperformance: the tendency to follow learning processing strategies of surface type rather than a deep type, the tendency with regard to learning regulation to study quite independent from the curriculum of the school, negative epistemic emotions that developed in high school mathematics classes, insufficient planning skills or lack of persistence, goal orientations that lack a competitive nature.

After linking underperformance with profiles of learning dispositions, the next stage is that of intervention. Intervention studies describe a rich arsenal of interventions linked to dispositional profiles, of which we will highlight some examples related to learning processing and regulation. For instance, Vermunt and Vermetten (2004) identified five "phenomena of dissonance in student learning patterns," with incompatibility of learning strategies and lack of integration between learning strategies being two examples. Especially in the situation of students transferring from one type of education to a new type, as in our context, Vermunt and Vermetten (2004) frequently observed such dissonant patterns. Their main solution, see Vermunt (2003), is in the adaptation of the teaching context by applying innovative teaching methods and practices. This was done in our context by providing a range of alternative formats and approaches for learners to comprehend mathematics in statistics, catering to different learning dispositions.

Other intervention studies see, e.g., Donche et al. (2012), enrich the range of intervention options by including different student feedback types. These researchers distinguish four different learning profiles, low and high self-efficacy in combination with low and high levels of learning regulation, and demonstrate that students of these profiles are characterised by different preferences for learning feedback and are thus best facilitated in their learning in different ways. For example, for a student like Maria, who was mostly watching videos on math problems after a class, understanding whether this comes from low self-efficacy/intense anxiety or procrastination/deficient self-regulation might fundamentally influence the type of support intervention. If it is the former, one could generate automatic feedback like "students like you Maria, who have expressed some anxiety in doing math have benefited tremendously by watching math video 34 before going to class. Therefore, it might be a good idea to consider watching this video 34 before class, but you may also watch it afterward." For the latter type of procrastination student one could provide automatic feedback like "student like you Maria who occasionally struggle to keep a clear study agenda, we recommend that you watch video 34 at least 3 hours before your next class meeting. This will help you to get well prepared for the class meeting."

Therefore, in precision education, our focus is not at the group level but the individual level. In the empirical part of our contribution, we focussed on the properties of four different profiles. The quartile split employed does no more that illustrate that differences in performance and differences in activity levels are associated with differences in learning dispositions. This happens at the individual level even stronger than at the group level: where in our analyses the mean levels of the four quarters may be indistinguishable for some of the dispositions, they may result in large variation at the individual level. It is that variation at the individual level that should serve as input for educational interventions.

Most learning dispositions of behavioural and cognitive type refer to study skills: think of planning, task management, learning processing, and regulation strategies. Every university has programs where these study skills are trained, often organised in extracurricular counselling classes. Therefore, making the step from diagnosis to intervention can be a small one. There are, however, two important differences between current counselling activities and DLA embedded interventions. The first is the sense of urgency: since the counselling is decoupled from the curriculum, it is difficult to demonstrate its relevance. The second difference relates to the generic nature of current counselling. The outcome of DLA is not only a disposition profile that gives students insights into their relative strengths and weakness but attached to that insight in the role these dispositions play in achieving good performance. That second insight will differ from module to module: dispositions relevant for learning mathematics will not be the same as dispositions important in, say, the study of languages. In this combination of these two types of information, DLA can bring its unique contribution to precision education.

The availability of such a broad range of disposition measurements as available in our study will be the exception rather than the rule. From that perspective, this study serves more as a showcase of what can be done with rich disposition data for precision education, where the way of getting such rich data may not be readily generalisable. An important facet of the richness of the data is having a full response of all students, where typically response rates of self-report surveys tend to be low and, typically, the missed cases represent students low in motivation and high in dropout risk, exactly those students it is crucial to have data about. A less crucial facet of the richness of data is the multitude of different disposition surveys. Disposition data tends to be collinear; that is, students with less favourable attitudes will tend to follow less adaptive learning strategies, or depend strongly on external types of motivation. The availability of specific interventions will govern in such a situation the choice of what type of survey instruments to apply: the ultimate goal of precision education is to prevent dropout rather than predict dropout.

References

Abdous, M., He, W., & Yen, C.-J. (2012). Using data mining for predicting relationships between online question theme and final grade. *Educational Technology & Society*, *15*(3), 77–88.

Buckingham Shum, S. & Deakin Crick, R. (2012). Learning dispositions and transferable competencies: Pedagogy, modelling and learning analytics. In S. Buckingham Shum, D. Gasevic, & R. Ferguson (Eds.), *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 92-101. New York, NY: ACM. doi:10.1145/2330601.2330629

Buckingham Shum, S. & Ferguson, R. (2012). Social learning analytics. *Educational Technology & Society*, *15*(3), 3-26.

Deakin Crick, R., Huang, S., Ahmed Shafi, A., & Goldspink, C. (2015). Developing resilient agency in learning: the internal structure of learning power. *British Journal of Educational Studies*, *63*(2), 121-160. doi:10.1080/00071005.2015.1006574

Donche, V., Coertjens, L., Vanthournout, G., & Van Petegem, P. (2012). Providing constructive feedback on learning patterns: An Individual learner's perspective. *Reflecting Education*, *8*(1), 114–131.

Elliot, A. J., Murayama, K., Kobeisy, A., & Lichtendfeld, S. (2015). Potential-based achievement goals. *British Journal of Educational Psychology*, *85*, 192-206. doi:10.1111/bjep.12051

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning, *TechTrends*, *59*(1), 64-71. doi:10.1007/s11528-014-0822-x

Gašević, D., Jovanović, J., Pardo, A., & Dawson, S. (2017). Detecting learning strategies with analytics: Links with self-reported measures and academic performance. *Journal of Earning Analytics*, *4*(1), 113–128. doi:10.18608/jla.2017.42.10

Hart, S. A. (2016). Precision education initiative: Moving toward personalized education, *Mind, Brain, and Education*. doi:10.1111/mbe.12109

Herodotou, C., Rienties, B., Hlosta, M., Boroowa, A., Mangafa, C., & Zdrahal, Z. (2020). The Scalable implementation of predictive learning analytics at a distance learning university: Insights from a longitudinal case study. *The Internet and Higher Education, 45*, 100725. doi:10.1016/j.iheduc.2020.100725

Hsiao, C. C., Huang, J. C. H., Huang, A. Y. Q., Lu, O. H. T., Yin, C. J. & Yang, S. J. H. (2019). Exploring the effects of online learning behaviors on short-term and long-term learning outcomes in flipped classrooms. *Interactive Learning Environments*, *27*(8), 1160-1177. doi:10.1080/10494820.2018.1522651

Huang, A. Y. Q., Lu, O. H. T., Huang, J. C. H., Yin C. J. & Yang, S. J. H. (2020). Predicting students' academic performance by using educational big data and learning analytics: evaluation of classification methods and learning logs. *Interactive Learning Environments*, *28*(2), 206-230. doi:10.1080/10494820.2019.1636086

Hwang, G.-J., Spikol, D., & Li, K.-C. (2018). Guest editorial: Trends and research issues of learning analytics and educational big data. *Educational Technology & Society*, *21*(2), 134–136.

Ifenthaler, D. (2015). Learning analytics. In J. M. Spector (Ed.), *Encyclopedia of Educational Technology* (Vol. 2, pp. 447–451). Thousand Oaks, CA: Sage.

Lian, A.-P. & Sangarun, P. (2017). Precision language education: A Glimpse into a possible future. GEMA Online® *Journal of Language Studies*, *17*(4). doi:10.17576/gema-2017-1704-01

Lin, C.-J., & Hwang, G.-J. (2018). A Learning analytics Approach to investigating factors affecting EFL students' oral performance in a flipped classroom. *Educational Technology & Society*, *21*(2), 205–219.

Lu, O. H. T., Huang, A. Y. Q., Lin, A. J. Q., Ogata, H., & Yang, S. J. H. (2018). Applying learning analytics for the early prediction of students' academic performance in blended learning. *Educational Technology & Society*, *21*(2), 220–232.

Martin, A. J. (2007). Examining a multidimensional model of student motivation and engagement using a construct validation approach. *British Journal of Educational Psychology*, *77*(2), 413-440. doi:10.1348/000709906X118036

Nguyen, Q., Tempelaar, D.T., Rienties, B., & Giesbers, B. (2016). What learning analytics based prediction models tell us about feedback preferences of students. *Quarterly Review of Distance Education*, *17*(3), 13-33.

Non, A. & Tempelaar, D. (2016). Time preferences, study effort, and academic performance. *Economics of Education Review*, *54*, 36–61. doi:10.1016/j.econedurev.2016.06.003

Padulo, L. & Arbib, M. A. (1974). *System theory*. Philadelphia: W. B. Saunders Company.

Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 259–282). doi:10.1007/978-1-4614-2018-7_12

Pekrun, R., Vogl, E., Muis, K. R., & Sinatra, G. M. (2017). Measuring emotions during epistemic activities: The Epistemically-Related Emotion Scales. *Cognition and Emotion*, *31*(6), 1268-1276. doi:10.1080/02699931.2016.1204989

Rienties, B., Cross, S., & Zdrahal, Z (2017). Implementing a learning analytics intervention and evaluation framework: What works? In B. Kei Daniel (Ed.), *Big data and learning analytics: Current theory and practice in higher education* (pp. 147–166). doi:10.1007/978-3-319-06520-5_10

Rienties, B., Tempelaar, D., Nguyen, Q., & Littlejohn, A. (2019). Unpacking the intertemporal impact of self-regulation in a blended mathematics environment. *Computers in Human Behavior*, *100*, 345-357. doi:10.1016/j.chb.2019.07.007

Siemens, G., & Gašević, D. (2012). Guest editorial - Learning and knowledge analytics. *Educational Technology & Society*, *15*(3), 1–2.

Tempelaar, D. T., Gijselaers, W. H., Schim van der Loeff, S., & Nijhuis, J. F. H., (2007). A Structural equation model analyzing the relationship of student achievement motivations and personality factors in a range of academic subject-matter areas. *Contemporary Educational Psychology*, *32*(1), 105-131. doi:10.1016/j.cedpsych.2006.10.004

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior, 47*, 157-167. doi:10.1016/j.chb.2014.05.038

Tempelaar, D. T., Rienties, B., Mittelmeier, J., & Nguyen, Q. (2018a). Student profiling in a dispositional learning analytics application using formative assessment. *Computers in Human Behavior*, *78*, 408-420. doi:10.1016/j.chb.2017.08.

Tempelaar, D. T., Rienties, B., & Nguyen, Q. (2017a). Towards actionable learning analytics using dispositions. *IEEE Transactions on Education*, *10*(1), 6-16. doi:10.1109/TLT.2017.2662679

Tempelaar, D.T., Rienties, B., & Nguyen, Q. (2017b). Adding dispositions to create pedagogy-based learning analytics. *Zeitschrift für Hochschulentwicklung, ZFHE*, *12*(1), 15-35.

Tempelaar, D., Rienties, B. & Nguyen, Q. (2018b). Investigating learning strategies in a dispositional learning analytics context: the case of worked examples. In *Proceedings of the International Conference on Learning Analytics and Knowledge (LAK'18)* (pp. 201-205). doi:10.1145/3170358.3170385

Tempelaar, D., Nguyen, Q., & Rienties, B. (2019). Learning feedback based on dispositional learning analytics. In M. Virvou, E. Alepis, G. Tsihrintzis, & L. Jain (Eds.), *Machine Learning Paradigms* (Vol. 158, pp. 69-89). doi:10.1007/978-3-030-13743-4_5

Vermunt, J. D. (1996). Metacognitive, cognitive and affective aspects of learning styles and strategies: A Phenomenographic analysis. *Higher Education*, *31*, 25–50. doi:10.1007/BF00129106

Vermunt, J. D. (2003). The Power of learning environments and the quality of student learning. In E. de Corte, L. Verschaffel, N. Entwistle, & J. Van Merriënboer (Eds.), *Advances in learning and instruction series. Powerful learning environments: Unravelling basic components and dimensions* (pp. 109–124). Bingley, UK: Pergamon/Elsevier Science Ltd.

Vermunt, J. D., & Vermetten, Y. J. (2004). Patterns in student learning: relationships between learning strategies, conceptions of learning, and learning orientations. *Educational Psychology Review*, *16*, 359–384. doi:10.1007/s10648-004-0005-y

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The Current landscape of learning analytics in higher education. *Computers in Human Behavior, 89*, 98-110. doi:10.1016/j.chb.2018.07.027

Williams, A., Sun, Z., Xie, K., Garcia, E., Ashby, I., Exter, M., Largent, D., Lu, P., Szafron, D., Ahmed, S., Onuczko, T., Smith, J., Tempelaar, D. T., Bitting, K. S., Olcott Marshall, A., Christensen, E., Tseng, H., & Walsh, J. (2017). Flipping STEM. In L. Santos Green, J. Banas, & R. Perkins (Eds.), *The Flipped College Classroom, Conceptualized and Re-Conceptualized, Part II* (pp. 149-186). Switzerland: Springer International Publishing. doi:10.1007/978-3-319-41855-1_8.

Yang, S. J. H. (2019). Precision education: New challenges for AI in education [conference keynote]. In *Proceedings of the 27th International Conference on Computers in Education (ICCE)* (pp. XXVII-XXVIII). Kenting, Taiwan: Asia-Pacific Society for Computers in Education (APSCE).

Lin, H.-C., Tu, Y.-F., Hwang, G.-J., & Huang, H. (2021). From Precision Education to Precision Medicine: Factors Affecting Medical Staff's Intention to Learn to Use AI Applications in Hospitals. *Educational Technology & Society*, *24* (1), 123-137.

From Precision Education to Precision Medicine: Factors Affecting Medical Staff's Intention to Learn to Use AI Applications in Hospitals

Hui-Chen Lin[1], Yun-Fang Tu[2], Gwo-Jen Hwang[3*] and Hsin Huang[4]
[1]Department of Neurology, Tri-Service General Hospital National Defense Medical Center, Taiwan // [2]Department of Library and Information Science, Research and Development Center for Physical Education, Health, and Information Technology, Fu Jen Catholic University, Taiwan // [3]Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan // [4]Department of Nursing, Tri-Service General Hospital National Defense Medical Center, Taiwan // ceciliatsgh@gmail.com // sandy0692@gmail.com // gjhwang.academic@gmail.com // jhshing1029@gmail.com
[*]Corresponding author

ABSTRACT: Precision medicine has become an essential issue in the medical community as the quality of medical care is being emphasized nowadays. The technological data analysis and predictions made by Artificial Intelligence (AI) technologies have assisted medical staff in designing personalized medicine for patients, making AI technologies an important path to precision medicine. During the implementation of the new emerging technology, medical staff's learning intentions will have a great influence on its effectiveness. With reference to the Technology Acceptance Model, this study explored medical staff's attitudes, intentions, and relevant influencing factors in relation to AI application learning. A total of 285 valid questionnaires were collected. Five major factors, perceived usefulness (PU), perceived ease of use (PEU), subjective norms (SN), attitude towards AI use (ATU), and behavioral intention (BI), were used for analyzing the AI learning of medical staff in a hospital. Based on the SEM analytical results and the research model, the four endogenous constructs of PU, PEU, SN, and ATU explained 37.4% of the changes in BI. In this model, SN and PEU were the determining factors of BI. The total effects of SN and PEU were 0.448 and 0.408 respectively, followed by PU, with a total effect of 0.244. As a result, the intentions of medical staff to learn to use AI applications to support precision medicine can be predicted by SN, PEU, PU, and ATU. Among them, subjective norms considering the influences of both supervisors and peers, such as encouragement, communication, and sharing, may assist precision education in supporting the learning attitudes and behavior regarding precision medicine. The research results can provide recommendations for examining medical staff's intention to use AI applications.

Keywords: Artificial intelligence, Subjective norms, Precision medicine, Precision education, Technology Acceptance Model

1. Introduction

The rapid advancement of computer technologies has provided new opportunities to facilitate medical services. Researchers have indicated that the ability of a computer to process a large amount of biological data and to calculate predicted models is becoming increasingly reliable, in particular for assisting doctors in making more accurate judgments (Alhashmi, Salloum, & Abdallah, 2018; Li, Hu, Li, & You, 2019). For instance, the advancement of AI (Artificial Intelligence) technologies enables computer systems to emulate medical experts' competences in analyzing, predicting and making judgments, as well as providing second opinions or support during the medical diagnosis process (Esteva et al., 2019; Williams et al., 2018). In the past, many studies have used computerized technologies to assist in medical diagnosis with positive results. For example, researchers have applied computerized technologies to radiology, pathology, and dermatology for image analysis. A combination of computerized calculation and the clinical diagnosis of doctors can greatly improve the accuracy and reliability of the diagnosis (Geras, Mann, & Moy, 2019; Nakata, 2019). Wearable devices can be used to assist and record the measurement of body health, and to collect physiological parameters (Yetisen, Martinez-Hurtado, Ünal, Khademhosseini, & Butt, 2018). The use of smartphone applications can provide a real-time risk assessment showing the possibility of having malignant melanoma (Chuchu et al., 2018). Applications of machine learning can assist doctors in improving the accuracy of cancer diagnosis and detection (Cruz & Wishart, 2006). Big data analysis and machine learning algorithms can be used to assist with clinical decision-making, successful-predictive surgical outcomes and medical treatment (Kanevsky et al., 2016; Senders et al., 2018).

The application of AI in clinical diagnosis has been gradually increasing. For example, it has been applied to improve the diagnostic accuracy of diabetic retinopathy (Poly et al., 2019); to have a fast test of ischemic stroke caused by large vessel occlusion (Murray, Unberath, Hager, & Hui, 2020); to improve the quality of fracture detection and its categorization (Langerhuizen et al., 2019); to improve the accuracy of valvular heart disease screening and congenital heart defects by AI auscultation (Thompson, Reinisch, Unterberger, & Schriefl, 2019); to assist the diagnosis and identification of liver masses (Azer, 2019) and mammography (Rodríguez-Ruiz et al., 2019); to assist the planning of Disease Risk Management (Marciniak, Kotas, Kamiński, & Ciota, 2014); different new emerging technologies, prescription, and treatments (Bassuk, Zheng, Li, Tsang, & Mahajan, 2016; Camarillo, Krummel, & Salisbury Jr, 2004); as well as to analyze the predicted treatment outcomes (Catto et al., 2003). Through using the technique of AI analysis, the big data of medical treatment provides precise data for making inferences. As a result, the effectiveness and accuracy of clinical diagnosis can be significantly increased. Such an application mode is of great help for improving the quality of medical treatments and ensuring the safety of patients (Hunter et al., 2012) as well as for implementing precision medicine, which emphasizes the importance of making precise analyses during the medical diagnosis process with the assistance of emerging technologies (Ho et al., 2020; Kosorok & Laber, 2019).

AI is becoming an important technology for precision medicine since it not only emulates the decision-making process of human experts, but can also make a detailed analysis and objective predictions based on a large set of data. From this perspective, it is important to train medical staff to employ AI applications to analyze medical data. Nursing staff would also need to be trained using a multidisciplinary approach to measure and analyze the critical factors of understanding precision medicine (Chen, Xie, Zou, & Hwang, 2020; Hwang, Sung, Chang, & Huang, 2020). Some studies have pointed out that, at this stage, active planning to cultivate AI professionals is an essential task in clinical education (Liao, Hsu, Chu, & Chu, 2015; Pepito & Locsin, 2019; Risling, 2017). Moreover, a study has shown that, in actual practice, the medical staff's understanding, attitudes, and behavioral intentions regarding AI applications are the key to determining whether AI technologies can support medical applications. Simultaneously, the promotion of AI applications to support precision medicine will be a great success if we understand the relevant factors that influence medical staff's learning of AI applications for medical treatment (Chiu & Tsai, 2014; Wu, Li, & Fu, 2011).

However, multiple factors influence medical staff's usage and learning of the technologies, for example, the personal beliefs and the expectations of peers, supervisors, and organizations (Alhashmi et al., 2019; Wu et al., 2011; Zhao, Ni, & Zhou, 2018). Several researchers have illustrated that subjective norms could be important determinants of medical staff's attitudes toward using technologies to learn or work; that is, subjective norms could directly influence the intention of medical staff to adopt technologies (Chiu & Tsai, 2014; Wang & Wang, 2009). Some previous studies have also reported the impacts of other factors that could affect medical staff's perceptions of using technologies; for example, Chiu and Tsai (2014) stated that when the social environment can encourage medical staff to adopt technologies for continuing learning, they will have more confidence in and positive attitudes towards using it. On the other hand, some studies have shown that medical staff's subjective norms would not significantly predict their behavioral intention to use technologies (Chiu, Tsai, & Chiang, 2013; Teo, Milutinović, & Zhou, 2016). As AI is an advanced technology, most medical staff may be unfamiliar with it; however, many medical institutes have started promoting AI in medical training or workplaces. Therefore, attention must be paid to medical staff's subjective norms when investigating the possibility of their use of AI applications to support medical applications (Ursavaş, Yalçın, & Bakır, 2019). More importantly, it is necessary to know the factors affecting their intention to learn AI applications. Those influential factors could be important parameters for developing adaptive or personalized training systems or approaches, which are the key issue in precision education (Hart, 2016). Therefore, this study aims to investigate the subjective norms as well as the learning perceptions, attitudes, and behavioral intentions of medical staff relating to the use of AI applications to support precision medicine, and the relationships among these factors in the workplace. The findings could be a good reference for those instructors and policymakers in medical schools or institutes.

2. Literature review and model development

2.1. Artificial intelligence and precision medicine

Artificial intelligence (AI) refers to the computer technologies that simulate human intelligence, such that computer systems are able to think and act like humans by making decisions and solving problems (Duan, Edwards, & Dwivedi, 2019; Simmons & Chappell, 1988). In the past decades, many AI applications have been reported by researchers, including in the areas of industrial design (Renzi, Leali, Cavazzuti, & Andrisano, 2014), smart buildings (Panchalingam & Chan, 2019), smart cars (Miles & Walker, 2006), factory automation (Özdemir & Hekim, 2018), medical diagnosis (Nakata, 2019; Park & Han, 2018) and education (Popenici & Kerr, 2017). Researchers have indicated several benefits of using AI technologies, such as improving the accuracy of decision making (Yang & Lin, 2019), enabling 24-hour service (Lilianira, Syah, Pusaka, & Ramdhani, 2020), and providing instant and personalized supports (Santos, 2019).

In the field of AI application, the issue of precision medicine has drawn the attention of researchers and practitioners in recent years (Collins & Varmus, 2015). Using personalized medicine as the foundation, precision medicine refers to the strategies used in treatments, such as targeted drug and cell therapy, through comparing the gene sequences and lifestyles of healthy people and patients after analyzing the computerized big data. According to the medical condition, precision medicine can provide the most suitable and precise treatments for each patient, or can effectively control diseases (Jameson & Longo, 2015). However, precise diagnosis and personalized medicine are the two main aspects of precision medicine to which AI has been applied. The purpose of precise diagnosis is to reduce the errors and to improve the accuracy of diagnosis; for instance, it provides reliable suggestions for diagnosis (Aerts, 2016; Giger, 2018) and predicts the possibility of a gene that causes cancer (Jamal-Hanjani et al., 2014; Nakagawa & Fujita, 2018). Regarding personalized medicine, it is used to not only conduct personal treatments based on patients' inherited genes, but also uses deep learning to improve the accuracy of drug development simulation and modeling, and hence shortens the development time and reduces the cost of developing new drugs (Bassuk et al., 2016). Several previous studies have reported the use of this approach in accelerating the development of cancer medicine (Denny, Van Driest, Wei, & Roden, 2018; Friedman, Letai, Fisher, & Flaherty, 2015). Furthermore, personalized medicine includes the use of an AI robot to operate more precise surgeries, resulting in a reduction in errors and cost (Camarillo et al., 2004).

Based on the above evidence, the use of AI and data analytics technologies to improve medical and health-care quality has been highly expected. However, some clinical professionals believe that there are potential uncertainties during the calculation of AI algorithms (Begoli, Bhattacharya, & Kusnezov, 2019). They are also concerned about the morals, social issues, and laws derived from AI medicine due to a lack of understanding (Cave & Dihal, 2019; Zou & Schiebinger, 2018). There is a trend of AI being introduced into medical care because of its positive impacts. However, it is necessary for medical staff who engage in AI-related work to attend training on the usage of AI equipment to carry out relevant inspections and treatments. The training content should not only include the fundamental understanding of AI, but also address the ethical norms and the standardized procedures of AI that practitioners should adhere to (Winfield, 2019). Furthermore, in the working field of assisting medical staff to learn and apply AI applications, the active cultivation of AI should be applied in precision medicine as it is one of the important tasks of medical education. The key to successful AI promotion is to understand the factors affecting medical staff's AI acceptance in precision medicine. Researchers have pointed out that the medical staff's behavior regarding technology adoption could be influenced by the usability and usefulness of the technologies (Chiu & Tsai, 2014; Chiu et al., 2013; Teo et al., 2016). When evaluating medical staff's behavioral intentions regarding new technology learning, the impact of important people's expectations around them should be considered (Fishbein & Ajzen, 1975; Chiu & Tsai, 2014). Therefore, the factors may be important variables to support the development of adaptive or personalized training systems or approaches (Hwang, Xie, Wah, & Gašević, 2020). Accordingly, this study proposed four factors from TAM (including perceived usefulness, perceived ease of use, attitude towards AI use, and behavioral intention), added a social factor (i.e., subjective norms), and explored the relationship between these factors.

2.2. Theoretical models related to the acceptance of new technologies

Several previous studies have reported that the Technology Acceptance Model (TAM), Theory of Reasoned Action (TRA), Theory of Planned Behavior (TPB), and Unified Theory of Acceptance and Use of Technology (UTAUT) belong to the field of common theories and models (Chau & Hu, 2001; Holden & Karsh, 2010). Numerous studies have indicated that TAM, one of the most commonly used models, can be used to explore the technological acceptance of users and to explain the attitudes and behavioral intention of medical staff when using new technologies (Hsu & Wu, 2017; Kowitlawakul, 2011; Strudwick, 2015; Zhang, Cocosila, & Archer, 2010). It is concentrated on influencing users' intention or actual use of technologies (Al-Emran, Mezhuyev, & Kamaludin, 2018; Davis, Bagozzi, & Warshaw, 1989; Legris, Ingham, & Collerette, 2003). Perceived usefulness (PU) and perceived ease of use (PEU) are the two factors used to measure users' perceptions. If users believe such technology is useful, they will maintain a positive attitude. Moreover, if they believe that such technology can assist them in completing tasks in a relaxing and effective manner, they will have a strong intention to use it (Teo, 2019; Wang & Wang, 2009). Attitude towards technology use (ATU) and behavioral intention (BI) are the remaining factors in the TAM. Users' perceptions (i.e., PU and PEU) can influence their attitudes and behaviors regarding that technology (Sánchez-Prieto, Olmos-Migueláñez, & García-Peñalvo, 2017).

Despite TAM having been used in a variety of studies, some researchers have added other factors to extend the model, with the aim of understanding users' acceptance behavior (Kwateng, Atiemo, & Appiah, 2018; Ursavaş et al., 2019). For instance, researchers have applied the extended Technology Acceptance Model to explore healthcare students' acceptance of using electronic health records (EHR) in nursing education; researchers have indicated that it is important to cultivate students' positive attitudes and increase their practical perceptions of EHR (Kowitlawakul, Chan, Pulcini, & Wang, 2015). Apart from that, subjective norms, the social-external variable used in the extended TAM research, can assist in understanding an individual's acceptance and usage of the new technology (Yu et al., 2009). With the development of technology, it has been a trend for AI to support precision medicine and a goal for its implementation in medical education. It also highlights the importance of medical staff supporting and promoting AI applications in precision medicine (Hsu & Wu, 2017; Hwang, 2014; Shorey et al., 2019).

2.3. Subjective norms

Subjective norms are defined as the situation in which individuals are subjected to social pressure while taking certain kinds of actions. They include social norms, others' expectations, or the pressure given by organizations (Fishbein & Ajzen,1975). Many studies have indicated that subjective norms have been associated with issues relevant to organizations and to individuals. They have, furthermore, illustrated their influences on users' intentions to use technology-supported services, and that will be one of the important factors affecting users' acceptance of those technologies (Venkatesh & Davis, 2000). In addition to directly influencing medical staff's learning attitudes, subjective norms have been shown to have influences on their intention to use such technology through affecting perceived usefulness (PU) and perceived ease of use (PEU) (Alhashmi et al., 2019; Yu et al., 2009). Researchers have pointed out that nursing staff's subjective norms (e.g., colleague support) could influence their attitudes toward adopting technology to support continued learning (Chiu et al., 2013). However, some researchers have pointed out that users' subjective norms may not always significantly influence the use of that technology (Azizi & Khatony, 2019). This study therefore aimed at exploring the relationships between medical staff's subjective norms, perceptions, attitudes, and intention to use in terms of learning AI applications to support precision medicine in the workplace (Ketikidis, Dimitrovski, Lazuras, & Bath, 2012; Ursavaş et al., 2019).

3. Research model and hypotheses

Referencing the TAM as the foundation, this study explored medical staff's attitudes and intention to learn to use AI applications to support precision medicine in their workplace, in particular investigating the five factors of perceived usefulness (PU), perceived ease of use (PEU), subjective norms (SN), attitude towards AI use (ATU), and behavioral intention (BI). Figure 1 shows the research model for this study.

According to the literature, the medical staff's perceived ease of use, usefulness, subjective norms, and attitudes towards adopting technologies for learning could have effects on their behavioral intentions; their perceived usefulness and perceived ease of use also influence their attitudes toward adopting AI applications (Chiu & Tsai, 2014; Chiu et al., 2013; Teo et al., 2016; Teo, 2019; Wang & Wang, 2009). In addition, medical staff's perceived ease of use of AI applications could affect their perceptions of usefulness, attitudes and behavior (Ursavaş et al., 2019; Wang & Wang, 2009). Therefore, the following research hypotheses are proposed in the present study:

H1: PU has a significant positive effect on BI.
H2: PEU has a significant positive effect on BI.
H3: ATU has a significant positive effect on BI.
H4: SN has a significant positive effect on BI.
H5: PEU has a significant negative effect on PU.
H6: PU has a significant positive effect on ATU.
H7: PEU has a significant negative effect on ATU.

In this study, subjective norms refer to the medical staff's perceptions of their significant others' opinions or suggestions that relate to their acceptance of adopting AI applications. Some researchers have identified SN as an important variable in explaining medical staff's attitudes toward technology adoption, which directly affects the intention to use (Chiu & Tsai, 2014; Ursavaş et al., 2019; Wang & Wang, 2009). Researchers have also pointed out that SN directly links to perceived usefulness, perceived ease of use, and attitudes (Chiu & Tsai, 2014; Chiu et al., 2013; Ursavaş et al., 2019). Accordingly, the following research hypotheses are proposed:

H8: SN has a significant positive effect on PU.
H9: SN has a significant positive effect on ATU.

H10: SN has a significant positive effect on PEU.

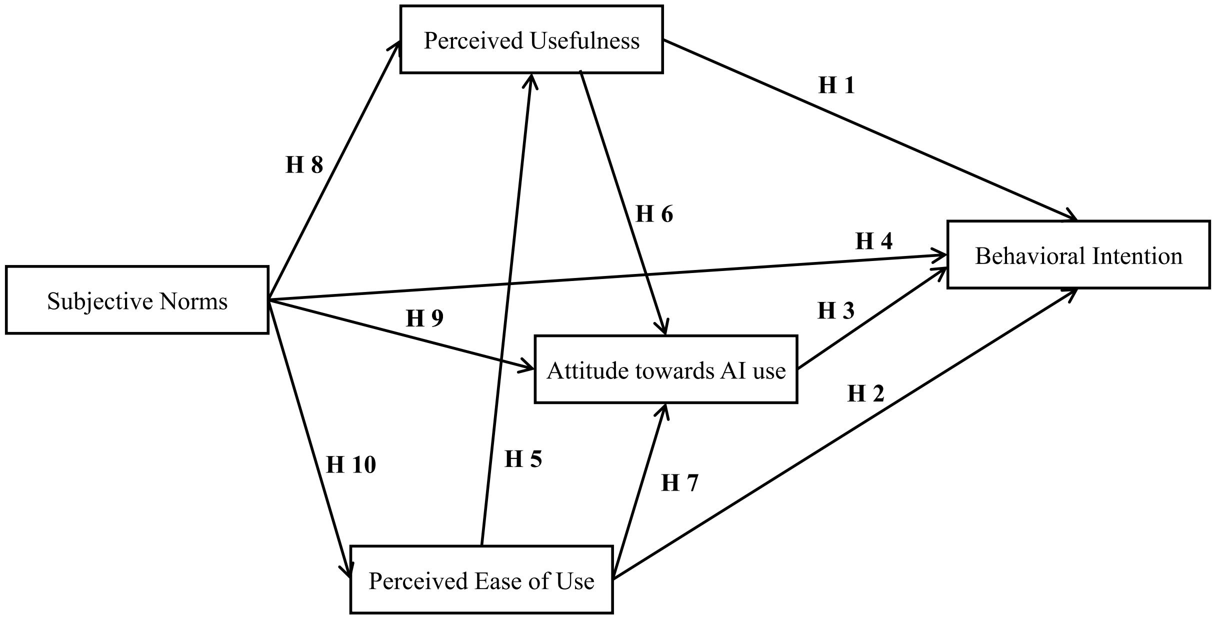


Figure 1. Proposed research model

4. Method

4.1. Participants

The participants of this study were medical staff working at a medical center in northern Taiwan. All participants were scheduled to complete two training modules, “AI and robotics in the New Health era” and “New era of medical education: AI-supported precision medicine” for 2 hours; following that, they were allowed to experience the use of an AI-based diagnosis system as shown in Figure 2 before completing the questionnaire. A total of 285 valid questionnaires were collected from 245 nursing staff and 40 physicians in this study.

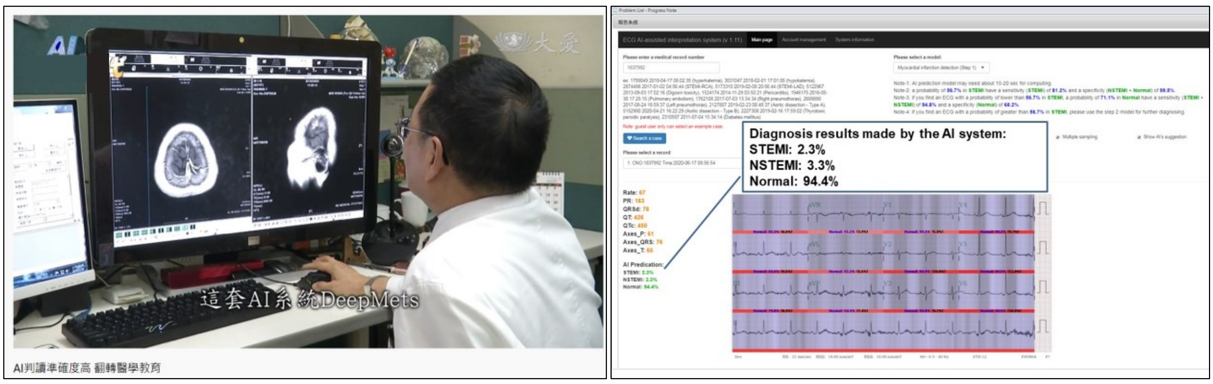


Figure 2. Learning materials and applications used in the training program

The demography of the sample is tabulated in Table 1. The gender distribution of participants was 13.3% male and 86.7% female. Regarding their age groups, 21-30 years old, 41-50 years old, and 31-40 years old constituted 30.2%, 29.8%, and 24.5% of total participants, respectively. Referring to their qualification levels, associate degrees, bachelor degrees, master degrees, and doctoral degrees constituted 58.2%, 23.5%, 15.4%, and 2.5%, respectively. Among them, the working experience of the medical staff was 2-5 years (29.1%), 6-10 years (21.4%), and above 10 years (49.5%).

All participants had experience of using the Internet, with 91.93% reporting that they used the Internet at least once a day. Furthermore, a majority of them reported that their average using time per day was approximately 3 to 5 hours (50.5%), 1 to 2 hours (32.6%), 0 to 1 hour (13%), and 6 hours and over (3.9%).

Table 1. Demography of the sample (N = 285)

Variable	Group	N	%
Sex	Female	247	86.7
	Male	38	13.3
Age	21-30 years	86	30.2
	31-40 years	70	24.5
	41-50 years	85	29.8
	51-60 years	35	12.3
	61 years and above	9	3.2
Educational qualification	Associate's degree	67	23.5
	Bachelor's degree	166	58.2
	Master's degree	44	15.4
	Doctoral degree	7	2.5
Working experience	2-5 years	83	29.1
	6-10 years	61	21.4
	11-15 years	34	11.9
	16-20 years	54	18.9
	21-25 years	24	8.4
	26 years and above	29	11.2
Frequency of using the Internet	at least once a day	262	91.9
	at least 3 times a week	22	7.7
	at least once a week	1	0.4
	less than once a week	0	0
Average time of using the Internet	0 to 1 hour per day	37	13.0
	1 to 2 hours per day	93	32.6
	3 to 5 hours per day	144	50.5
	6 hours and over per day	11	3.9

4.2. Instruments

The present study was based on Davis's study (1989) as the foundation and applied his scale items, which were adapted from published sources that reported a high degree of reliability (Chiu & Tsai, 2014; Teo & Zhou, 2014; Ursavaş et al., 2019; Wu et al., 2011). Four professionals were consulted during this study. Two are professors specializing in medical education and the other two are experts in scientific and technological education. The aim was to confirm that all items listed in the questionnaire could be used to completely understand medical staff's attitudes and intentions to learn to use AI applications to support precision medicine.

The instrument consists of participants' demographic information and 20 items, aiming at balancing their beliefs in five constructs with four items each. In terms of PU, participants will say "I believe that learning to use the AI-technology tools can better assist healthcare work"; in terms of PEU, they will say "Learning to use the AI-technology tools for healthcare is easy for me"; referring to the SN, they will mention "My supervisor or organization believes that I should employ the AI-technology tools to assist my healthcare work in the future"; referring to the ATU, they will mention "I have a generally favorable attitude toward learning the AI-technology tools"; referring to the BI, they will mention "I intend to learn the AI-technology tools for my healthcare work in the future."

The questionnaire used in this study applied a 5-point Likert scale, where 5 refers to *strongly agree* and 1 refers to *strongly disagree*. The preliminary analysis indicated that four items (i.e., PU04, ATU03, ATU04, and BI04) had low factor loadings or had a higher correlation with other items used in the model. These items were, therefore, deleted from further analysis. Eventually, 16 items were selected for the subsequent analysis (Appendix A). The final structure showed an excellent internal consistency and reliability, with alpha values ranging from .819 to .922, as shown in Table 2.

4.3. Data analysis

The present study employed AMOS in SPSS for the analysis. Firstly, the descriptive statistics were conducted to verify the skewness and kurtosis of values and to establish the univariate normality of the data. The critical values were ±3.0 and ±10.0, respectively (Kline, 2010). Furthermore, we tested the multivariate normality using

Mardia's normalized multivariate kurtosis (Mardia, 1970). The structure of the questionnaire was, thereafter, checked by confirmatory factor analysis (CFA) and the proposed hypotheses were verified, with the aim of exploring the relationships between PU, PEU, SN, ATU, and BI, in particular, influencing medical staff's learning to use AI applications to support precision medicine.

5. Results

5.1. Descriptive statistics

In this study, the means of the other constructs were between 3.838 and 3.977, with standard deviations between 0.555 and 0.724. The values of the skewness and kurtosis for the items were between -0.568 and 0.17, and -0.839 and 0.535, respectively, indicating univariate normality in the data (Kline, 2010). In this study, Mardia's coefficient was 68.548. According to the suggestion given by Bollen (1989), a multivariate normality will occur if Mardia's coefficient is less than p (p +2), where p refers to the number of observed variables. This study used 16 observed variables, and Mardia's coefficient was less than 288, indicating that the data had a multivariate normal distribution.

5.2. Test of the measurement model

The present study adopted the CFA as the measuring model. The estimation of overall model fit was made by $\chi 2$ and other fit indices, including the Tucker-Lewis index (TLI), the comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). Hu and Bentler (1999) indicated that the TLI and CFI show a good model fit if their statistics are greater than 0.95. They reported that RMSEA and SRMR values less than .06 and .08, respectively, are acceptable. From the results, the measurement model displayed an acceptable fit to the sample data (χ^2 = 194.48; χ^2/df = 1.870; TLI = .967; CFI = .975; RMSEA = .053; SRMR = .037).

Table 2. Results of the Confirmatory Factor Analysis

Items	UE	*t*-value*	SE	CR	AVE	Alpha value	Mean	*SD*
PU				0.893	0.736	0.892	3.923	0.724
PU01#	1		0.901					
PU02	0.953	19.278	0.871					
PU03	0.924	16.958	0.798					
PEU				0.849	0.585	0.845	3.928	0.600
PEU01#	1		0.825					
PEU02	0.957	12.125	0.721					
PEU03	0.934	12.316	0.736					
PEU04	0.869	14.115	0.772					
SN				0.924	0.754	0.922	3.838	0.710
SN01#	1		0.894					
SN02	0.872	17.295	0.8					
SN03	0.937	20.757	0.879					
SN04	0.864	22.546	0.896					
ATU				0.820	0.695	0.819	3.977	0.656
ATU01	1		0.848					
ATU02#	1.044	13.196	0.819					
BI				0.827	0.616	0.824	3.860	0.555
BI01#	1		0.694					
BI02	1.042	11.937	0.806					
BI03	1.252	10.65	0.847					

Note. UE = unstandardized estimate; SE = standardized estimate, factor loadings; SN = subjective norms; PU = perceived usefulness; PEU = perceived ease of use; ATU = attitude towards AI use; BI = behavioral intention.
* p < .01; # this value was fixed at 1.000 for model identification purposes.

Table 2 describes the CFA result; all the factor loadings of the measured items are higher than the threshold value of 0.60 (ranging from 0.694 to 0.901). The values of Cronbach's alpha of PU, PEU, SN, ATU, and BI were .892, .845, .922, .819, and .824, respectively. The overall reliability of the questionnaire was .912, indicating a sufficient internal consistency of the factor items. Moreover, the ranges of composite reliability (CR) were between 0.820 and 0.924, and the ranges of average variance extracted (AVE) were between 0.585 and 0.736, indicating that the present study had a good convergence validity of the adopted variables. Therefore, the convergence validity of all of the variables used in this study was confirmed (Fornell & Larcker, 1981).

Apart from the convergence validity, the square roots of the AVE of all variables used were greater than the correlation coefficient. Therefore, the variables used in this study have different validities (Farrell, 2010), as shown in Table 3.

Table 3. Correlation coefficient and discriminant validity

	PU	PEU	SN	ATU	BI
PU	(0.858)				
PEU	0.489	(0.765)			
SN	0.475	0.486	(0.868)		
ATU	0.754	0.490	0.489	(0.834)	
BI	0.486	0.530	0.448	0.472	(0.785)

Note. The diagonal value is the square root of AVE, the construct; SN= subjective norms; PU= perceived usefulness; PEU= perceived ease of use; ATU= attitude towards AI use; BI= behavioral intention.

5.3. Tests of direct and indirect effects

The results of the structural model showed a good model ($\chi^2 = 200.358$; $\chi^2/df = 2.131$; TLI = 0.952; CFI = 0.963; RMSEA = 0.063; SRMR = .044). Based on the hypotheses proposed in this study, the bootstrap method was performed for the evaluation. As shown in Table 4, six out of 10 hypotheses were supported by the data; except for H1, H3, H7, and H9, all other hypotheses were supported in this study (see Figure 3).

Table 4. Results of Hypothesis Testing.

Hypotheses	Path	Estimate	*t*-value	Bias-corrected		Sig	Result
				Lower	Upper	*p*	
H1	PU→BI	0.175	1.645	-0.063	0.418	0.136	Not supported
H2	PEU→BI	0.313	3.916	0.143	0.458	0.001	Supported
H3	ATU→BI	0.108	0.953	-0.161	0.357	0.395	Not supported
H4	SN→BI	0.161	2.225	0.014	0.305	0.032	Supported
H5	PEU→PU	0.339	4.73	0.189	0.476	0.001	Supported
H6	PU→ATU	0.636	9.179	0.503	0.755	0.001	Supported
H7	PEU→ATU	0.115	1.715	-0.017	0.251	0.09	Not supported
H8	SN→PU	0.31	4.599	0.174	0.445	0.001	Supported
H9	SN→ATU	0.131	2.087	-0.012	0.254	0.081	Not supported
H10	SN→PEU	0.486	7.575	0.366	0.584	0.001	Supported

Note. PU = perceived usefulness; PEU = perceived ease of use; SN = subjective norms; ATU = attitude towards AI use; BI = behavioral intention.

Table 5 shows the standardized total effects, direct and indirect effects among each variable in the model. The addition of the direct effects and the indirect effects is equal to the total effects. In the model used in this study, the standardized total effects of predictor variables on the dependent variables ranged from 0.108 to 0.636.

According to the research model, four endogenous constructs were tested. The coefficient of variation of BI was determined by PU, PEU, SN, and ATU, and the explanatory power (R2) was 0.374. The changes of BI (37.4%) were explained by PU, PEU, SN, and ATU. Regarding the variations of other endogenous constructs, PU (31.3%), PEU (23.6%), and ATU (60%) were explained by their determinants.

Regarding these four endogenous constructs, the highest amount of variance (60%) was explained by the determinants of ATU. The most dominant determinant was PU and its total effect was 0.636. The second dominant determinant was SN and its total effect was 0.489. The total effect of PEU was 0.331. The explained variation of BI in this model was 0.374. It was mainly determined by SN and PEU, and their total effects were 0.448 and 0.408. Following this, the total effect of PU was 0.244. However, the total effect of ATU on BI was 0.108 and it was a statistically insignificant effect. The explained variation of PU was 0.313 and it was mainly determined by SN and PEU as their total effects were 0.474 and 0.339, respectively. The explained variation of PEU was 0.236, and it was mainly determined by SN with 0.486 as the total effect.

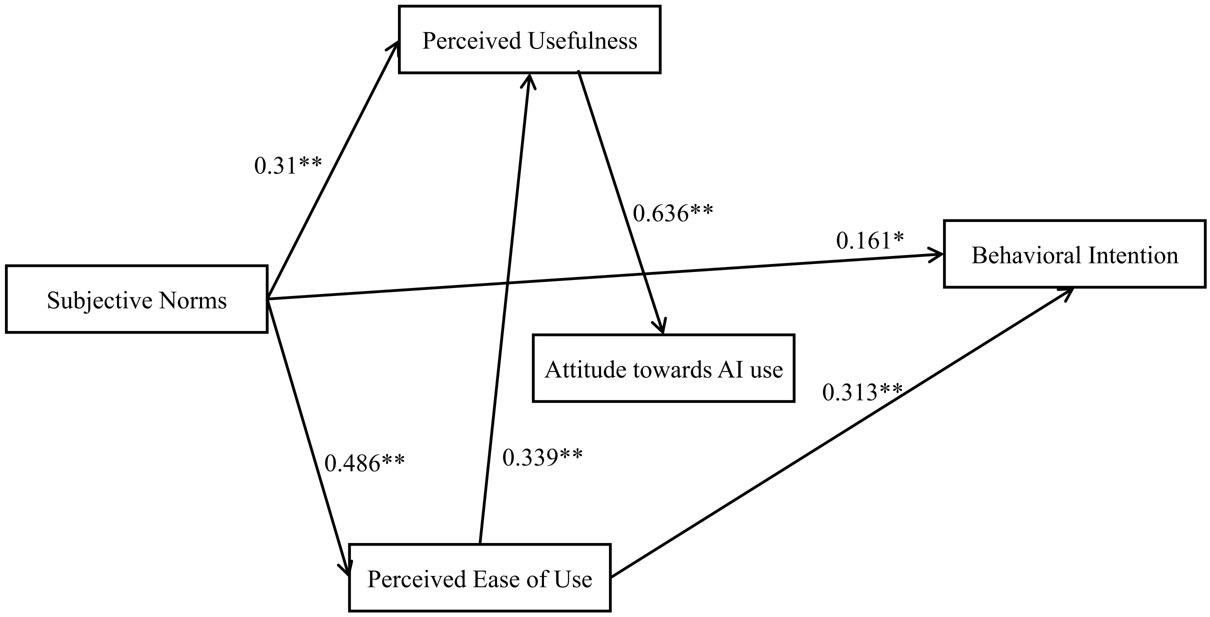


Figure 3. Results of the research model

Table 5. Direct, indirect and total effects of the research model

Endogenous variable	Determinant	Standardized estimates		
		Direct	Indirect	Total
PU (R^2 = 0.313)	PEU	0.339	-	0.339
	SN	0.310	0.164	0.474
PEU (R^2 = 0.236)	SN	0.486	-	0.486
ATU (R^2 = 0.600)	PU	0.636	-	0.636
	PEU	0.115	0.215	0.331
	SN	0.131	0.358	0.489
BI (R^2 = 0.374)	PU	0.175	0.069	0.244
	PEU	0.313	0.095	0.408
	SN	0.161	0.288	0.448
	ATU	0.108	-	0.108

Note. PU = perceived usefulness; PEU = perceived ease of use; SN = subjective norms; ATU = attitude towards AI use; BI = behavioral intention.

6. Discussion and conclusions

The present study examined the relationship among subjective norms and PU, PEU, ATU and BI when medical staff learn to employ AI applications to support precision medicine. Firstly, this study adopted CFA to establish a five-factor structure. Based on the analytical results, this investigation was effective and reliable. Moreover, there were predictive relationships between PU, PEU, SN, ATU, and BI. The findings are consistent with the results of previous studies relating to medical staff's learning attitudes, use intention, and perceptions of technologies (Chiu & Tsai, 2014; Teo, 2019; Ursavaş et al., 2019; Wu et al., 2011). This study then adopted SEM for testing the proposed hypotheses, showing that medical staff's perceived usefulness of learning AI applications to support precision medicine would affect their learning attitude (H6). The medical staff's perceived ease of use of learning about AI applications to support precision medicine would affect their perceived usefulness (H5), which would also directly influence their behavioral intentions (H2). In other words, if the technology is not easy to operate, even if it is useful to users, they may remain in their original situation or choose other options (Teo, 2019).

Moreover, this study also found that medical staff's SN could predict their perceived usefulness (H8), perceived ease of use (H10), and behavioral intention (H4) to learn to use AI applications to support precision medicine. As SN represents the person or organization that has the power to determine and support specified events, it is important and interesting to know the role of SN in AI education (Li, Liu, & Rojas-Méndez, 2013). In particular, the medical staff generally need to work in teams. In such a team-working culture, they tend to accept the instructions or requests from the person at the management level in order to achieve the goal of the team. This could be the reason why subjective norms significantly influence the medical staff's behavior intention of adopting and learning AI applications (Chiu & Tsai, 2014; Chiu et al., 2013; Schmidt & Diestel, 2011). Some researchers have also indicated that subjective norms could be determined by the perceived pressure from social views or regulations to influence people's behaviors and manners to comply with the views or regulations (Ursavaş et al., 2019). In medical working environments, medical staff are trained to strictly follow the regulations in each step of the medical diagnosis and treatment process since failing to follow those regulations could endanger the patients (Chiu & Tsai, 2014).

Some of the hypotheses in this study are not significant. For instance, SN did not have a significant influence on ATU (H9). As medical staff generally work in a particular environment, which has a "problem orientation," they cooperate as a team to identify and solve problems. Therefore, SN would not be the main factor affecting their attitudes toward learning or working, although it would determine their behavioral intention. Similarly, ATU would not decide their behavioral intention either (H3). Medical staff might have their own attitudes toward learning AI applications; however, SN or missions generally outweigh their attitudes when making decisions related to their work. This finding is evidenced by several researchers who indicated the importance of a supportive organizational climate (i.e., the common value in an organization) to medical staff (Chiu & Tsai, 2014; Chiu et al., 2013; Schmidt & Diestel, 2011). Another important finding of the present study is that medical staff's subjective norms can further influence their attitudes towards learning to use AI applications through their perception of how AI applications can better assist healthcare. This also echoes the point that medical staff generally value the usefulness of a new technology to their work. If subjective norms deliver correct information to help them understand the usefulness of the new technology, they could change their attitudes toward learning or using it.

To sum up, subjective norms are an important factor influencing the adoption of AI applications in medical institutes for supporting precision medicine. In other words, medical institutes should consider the influences of supervisors and peers on medical staff. Positive opinions, for example encouragement, communication, and sharing, given by supervisors and peers can strengthen the expectations and confidence of medical staff. In turn, they may influence their perceptions and use intention regarding AI applications to support precision medicine (Chiu & Tsai, 2014; Zhao et al., 2018). As encouraging medical staff to learn AI applications is the basis for implementing precision medicine, based on the findings of this study, helping decision makers or management level staff to know the importance of promoting AI applications in their institutes is very important. Therefore, it is recommended that when trying to promote precision medicine, it is necessary to have a workshop or training program for those management level staff or relevant policymakers in the medical institutes. In the meantime, from the perspective of precision education, the findings of the present study could be a reference for those who intend to implement training programs for AI applications for medical staff. It has been found that, in addition to subjective norms, perceived ease of use is an important factor affecting medical staff's behavior intention. This implies that the development of adaptive learning systems needs to be considered from the perspective of the user interface in addition to the learning content or learning paths. As indicated by several previous studies, a proper user-interface design which takes into account individual learners' needs could significantly affect learning performances (Yang, Hwang, & Yang, 2013).

This study has some limitations. Regarding the samples, it focuses on medical staff from Taiwan, limiting the research inference. It is suggested that larger samples be used to explore the attitudes and behaviors of medical staff from different areas regarding learning to use AI applications to support precision medicine. It is also recommended that some external variables be considered when exploring their attitudes and behaviors. For example, facilitating conditions, anxiety, self-efficacy, training, and job relevance can all be considered as external variables. In the future, intervention experiments and interviews can be designed to investigate the teaching modes referencing the AI environment. They could provide a deeper understanding of medical staff's attitudes and explore relevant influencing factors and effectiveness in learning to use AI applications to support precision medicine.

Acknowledgement

This study is supported in part by the Ministry of Science and Technology of the Republic of China under contract numbers MOST-109-2511-H-011-002-MY3 and MOST-108-2511-H-011-005-MY3.

References

Aerts, H. J. (2016). The Potential of radiomic-based phenotyping in precision medicine: A Review. *JAMA Oncology, 2*(12), 1636-1642.

Al-Emran, M., Mezhuyev, V., & Kamaludin, A. (2018). Technology acceptance model in m-learning context: A Systematic review. *Computers & Education*, *125*, 389-412.

Alhashmi, S. F., Salloum, S. A., & Abdallah, S. (2019). Critical success factors for implementing artificial intelligence (AI) projects in Dubai government United Arab Emirates (UAE) health sector: Applying the extended technology acceptance model (TAM). In *International Conference on Advanced Intelligent Systems and Informatics* (pp. 393-405). doi:10.1007/978-3-030-31129-2_36

Azer, S. A. (2019). Deep learning with convolutional neural networks for identification of liver masses and hepatocellular carcinoma: A Systematic review. *World Journal of Gastrointestinal Oncology, 11*(12), 1218-1230.

Azizi, S. M., & Khatony, A. (2019). Investigating factors affecting on medical sciences students' intention to adopt mobile learning. *BMC Medical Education, 19*(1), 381. doi:10.1186/s12909-019-1831-4

Bassuk, A. G., Zheng, A., Li, Y., Tsang, S. H., & Mahajan, V. B. (2016). Precision medicine: genetic repair of retinitis pigmentosa in patient-derived stem cells. *Scientific Reports, 6*, 19969. doi:10.1038/srep19969

Begoli, E., Bhattacharya, T., & Kusnezov, D. (2019). The Need for uncertainty quantification in machine-assisted medical decision making. *Nature Machine Intelligence, 1*(1), 20-23.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: John Wiley and Sons, Inc.

Camarillo, D. B., Krummel, T. M., & Salisbury Jr, J. K. (2004). Robotic technology in surgery: Past, present, and future. *The American Journal of Surgery, 188*(4), 2-15.

Catto, J. W., Linkens, D. A., Abbod, M. F., Chen, M., Burton, J. L., Feeley, K. M., & Hamdy, F. C. (2003). Artificial intelligence in predicting bladder cancer outcome: A Comparison of neuro-fuzzy modeling and artificial neural networks. *Clinical Cancer Research, 9*(11), 4172-4177.

Cave, S., & Dihal, K. (2019). Hopes and fears for intelligent machines in fiction and reality. *Nature Machine Intelligence, 1*(2), 74-78.

Chau, P. Y., & Hu, P. J. H. (2001). Information technology acceptance by individual professionals: A Model comparison approach. *Decision Sciences, 32*(4), 699-719.

Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020). Application and theory gaps during the rise of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence, 1*, 100002. doi:10.1016/j.caeai.2020.100002

Chiu, Y. L., & Tsai, C. C. (2014). The Roles of social factor and internet self-efficacy in nurses' web-based continuing learning. *Nurse Education Today, 34*(3), 446-450.

Chiu, Y. L., Tsai, C. C., & Chiang, C. Y. F. (2013). The Relationships among nurses' job characteristics and attitudes toward web-based continuing learning. *Nurse education today, 33*(4), 327-333.

Chuchu, N., Takwoingi, Y., Dinnes, J., Matin, R. N., Bassett, O., Moreau, J. F., Bayliss, S. E., Davenport, C., Godfrey, K., O'Connell, S., Jian, A., Walter, F. M., & Deeks, J. J. (2018). Smartphone applications for triaging adults with skin lesions that are suspicious for melanoma. *Cochrane Database of Systematic Reviews*(12), CD013192.

Collins, F. S., & Varmus, H. (2015). A New initiative on precision medicine. *New England Journal of Medicine, 372*(9), 793-795.

Cruz, J. A., & Wishart, D. S. (2006). Applications of machine learning in cancer prediction and prognosis. *Cancer Informatics, 2*, 59-78.

Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A Comparison of two theoretical models. *Management Science, 35*(8), 982-1003.

Denny, J. C., Van Driest, S. L., Wei, W. Q., & Roden, D. M. (2018). The Influence of big (clinical) data and genomics on precision medicine and drug development. *Clinical Pharmacology & Therapeutics, 103*(3), 409-418.

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. *International Journal of Information Management, 48,* 63-71.

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A Guide to deep learning in healthcare. *Nature Medicine, 25*(1), 24-29.

Farrell, A. M. (2010). Insufficient discriminant validity: A Comment on Bove, Pervan, Beatty, and Shiu (2009). *Journal of Business Research, 63*(3), 324-327.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An Introduction to theory and research*. Reading, MA: Addison-Wesley.

Fornell, C., & Larcker, D. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50.

Friedman, A. A., Letai, A., Fisher, D. E., & Flaherty, K. T. (2015). Precision medicine for cancer with next-generation functional diagnostics. *Nature Reviews Cancer, 15*(12), 747-756.

Geras, K. J., Mann, R. M., & Moy, L. (2019). Artificial intelligence for mammography and digital breast tomosynthesis: current concepts and future perspectives. *Radiology, 293*(2), 246-259.

Giger, M. L. (2018). Machine learning in medical imaging. *Journal of the American College of Radiology, 15*(3), 512-520.

Hart, S. A. (2016). Precision education initiative: moving toward personalized education. *Mind, Brain, and Education, 10*(4), 209-211.

Ho, D., Quake, S. R., McCabe, E. R., Chng, W. J., Chow, E. K., Ding, X., … Zarrinpar, A. (2020). Enabling technologies for personalized and precision medicine. *Trends in Biotechnology, 38*(5), 497-518.

Holden, R. J., & Karsh, B. T. (2010). The Technology acceptance model: Its past and its future in health care. *Journal of Biomedical Informatics, 43*(1), 159-172.

Hsu, H. H., & Wu, Y. H. (2017). Investigation of the effects of a nursing information system by using the technology acceptance model. *CIN: Computers, Informatics, Nursing, 35*(6), 315-322.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Hunter, J., Freer, Y., Gatt, A., Reiter, E., Sripada, S., & Sykes, C. (2012). Automatic generation of natural language nursing shift summaries in neonatal intensive care: BT-Nurse. *Artificial Intelligence in Medicine, 56*(3), 157-172.

Hwang, G. J. (2014). Definition, framework and research issues of smart learning environments-a context-aware ubiquitous learning perspective. *Smart Learning Environments, 1*(1), 4. doi:10.1186/s40561-014-0004-5

Hwang, G. J., Sung, H. Y., Chang, S. C., & Huang, X. C. (2020). A Fuzzy expert system-based adaptive learning approach to improving students' learning performances by considering affective and cognitive factors. *Computers and Education: Artificial Intelligence, 1*, 00003. doi:10.1016/j.caeai.2020.100003

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers & Education: Artificial Intelligence, 1*, 100001. doi:10.1016/j.caeai.2020.100001

Jamal-Hanjani, M., Hackshaw, A., Ngai, Y., Shaw, J., Dive, C., Quezada, S., … Swanton, C. (2014). Tracking genomic cancer evolution for precision medicine: The Lung TRACERx study. *PLoS Biology, 12*(7), e1001906. doi:10.1371/journal.pbio.1001906

Jameson, J. L., & Longo, D. L. (2015). Precision medicine—Personalized, problematic, and promising. *Obstetrical & gynecological Survey, 70*(10), 612-614.

Kanevsky, J., Corban, J., Gaster, R., Kanevsky, A., Lin, S., & Gilardino, M. (2016). Big data and machine learning in plastic surgery: A New frontier in surgical innovation. *Plastic and Reconstructive Surgery, 137*(5), 890e-897e. doi:10.1097/PRS.0000000000002088

Ketikidis, P., Dimitrovski, T., Lazuras, L., & Bath, P. A. (2012). Acceptance of health information technology in health professionals: An Application of the revised technology acceptance model. *Health Informatics Journal, 18*(2), 124-134.

Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.

Kosorok, M. R., & Laber, E. B. (2019). Precision medicine. *Annual Review of Statistics and Its Application, 6*, 263-286.

Kowitlawakul, Y. (2011). The Technology acceptance model: Predicting nurses' intention to use telemedicine technology (eICU). *CIN: Computers, Informatics, Nursing, 29*(7), 411-418.

Kowitlawakul, Y., Chan, S. W. C., Pulcini, J., & Wang, W. (2015). Factors influencing nursing students' acceptance of electronic health records for nursing education (EHRNE) software program. *Nurse Education Today, 35*(1), 189-194.

Kwateng, K. O., Atiemo, K. A. O., & Appiah, C. (2018). Acceptance and use of mobile banking: An Application of UTAUT2. *Journal of Enterprise Information Management, 32*(1), 118-151.

Langerhuizen, D. W., Janssen, S. J., Mallee, W. H., van den Bekerom, M. P., Ring, D., Kerkhoffs, G. M., … Doornberg, J. N. (2019). What are the applications and limitations of artificial intelligence for fracture detection and classification in orthopaedic trauma imaging? A Systematic review. *Clinical Orthopaedics and Related Research®, 477*(11), 2482-2491. doi:10.1097/CORR.0000000000000848

Legris, P., Ingham, J., & Collerette, P. (2003). Why do people use information technology? A Critical review of the technology acceptance model. *Information & Management, 40*(3), 191-204.

Li, J., Liu, F., & Rojas-Méndez, J. I. (2013). How international students select offshore programs: The Influence of image, attitude, subject norm, and perceived behavioral control. *Asia Pacific Education Review, 14*(3), 381-390.

Li, X., Hu, B., Li, H., & You, B. (2019). Application of artificial intelligence in the diagnosis of multiple primary lung cancer. *Thoracic Cancer, 10*(11), 2168-2174.

Liao, P. H., Hsu, P. T., Chu, W., & Chu, W. C. (2015). Applying artificial intelligence technology to support decision-making in nursing: A case study in Taiwan. *Health Informatics Journal, 21*(2), 137-148.

Lilianira, F. T., Syah, T. Y. R., Pusaka, S., & Ramdhani, D. (2020). The Technology processing of "4 care" home care application based on artificial intelligence. *Journal of Multidisciplinary Academic, 4*(1), 6-9.

Marciniak, P., Kotas, R., Kamiński, M., & Ciota, Z. (2014). Implementation of artificial intelligence methods on example of cardiovascular diseases risk stratification. In *Proceedings of the 21st International Conference Mixed Design of Integrated Circuits and Systems (MIXDES)* (pp. 503-507). doi:10.1109/MIXDES.2014.6872251

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*(3), 519-530.

Miles, J., & Walker, A. (2006). The Potential application of artificial intelligence in transport. *IEE Proceedings-Intelligent Transport Systems, 153*(3), 183-198. doi:10.1049/ip-its:20060014

Murray, N. M., Unberath, M., Hager, G. D., & Hui, F. K. (2020). Artificial intelligence to diagnose ischemic stroke and identify large vessel occlusions: A Systematic review. *Journal of NeuroInterventional Surgery, 12*(2), 156-164.

Nakagawa, H., & Fujita, M. (2018). Whole genome sequencing analysis for cancer genomics and precision medicine. *Cancer Science, 109*(3), 513-522.

Nakata, N. (2019). Recent technical development of artificial intelligence for diagnostic medical imaging. *Japanese journal of radiology, 37*(2), 103-108.

Özdemir, V., & Hekim, N. (2018). Birth of industry 5.0: Making sense of big data with artificial intelligence,"the internet of things" and next-generation technology policy. *Omics: A Journal of Integrative Biology, 22*(1), 65-76.

Panchalingam, R., & Chan, K. C. (2019). A state-of-the-art review on artificial intelligence for smart buildings. *Intelligent Buildings International*, 1-24.

Park, S. H., & Han, K. (2018). Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology, 286*(3), 800-809.

Pepito, J. A., & Locsin, R. (2019). Can nurses remain relevant in a technologically advanced future?. *International Journal of Nursing Sciences, 6*(1), 106-110.

Poly, T. N., Islam, M. M., Yang, H. C., Nguyen, P. A., Wu, C. C., & Li, Y. C. J. (2019). Artificial intelligence in diabetic retinopathy: Insights from a meta-analysis of deep learning. *Studies in Health Technology and Informatics, 264*, 1556-1557.

Popenici, S. A., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning, 12*(1), 22.

Renzi, C., Leali, F., Cavazzuti, M., & Andrisano, A. O. (2014). A review on artificial intelligence applications to the optimal design of dedicated and reconfigurable manufacturing systems. *The International Journal of Advanced Manufacturing Technology, 72*(1-4), 403-418.

Risling, T. (2017). Educating the nurses of 2025: Technology trends of the next decade. *Nurse Education in Practice, 22*, 89-92.

Rodríguez-Ruiz, A., Krupinski, E., Mordang, J. J., Schilling, K., Heywang-Köbrunner, S. H., Sechopoulos, I., & Mann, R. M. (2019). Detection of breast cancer with mammography: Effect of an artificial intelligence support system. *Radiology, 290*(2), 305-314.

Sánchez-Prieto, J. C., Olmos-Migueláñez, S., & García-Peñalvo, F. J. (2017). MLearning and pre-service teachers: An Assessment of the behavioral intention using an expanded TAM model. *Computers in Human Behavior, 72*, 644-654.

Santos, O. C. (2019). Artificial intelligence in psychomotor learning: Modeling human motion from inertial sensor data. *International Journal on Artificial Intelligence Tools, 28*(04), 1940006. doi:10.1142/S0218213019400062

Schmidt, K. H., & Diestel, S. (2011). Differential effects of decision latitude and control on the job demands–strain relationship: A Cross-sectional survey study among elderly care nursing staff. *International Journal of Nursing Studies, 48*(3), 307-317.

Senders, J. T., Staples, P. C., Karhade, A. V., Zaki, M. M., Gormley, W. B., Broekman, M. L. D., Smith, T. R., & Arnaout, O. (2018). Machine learning and neurosurgical outcome prediction: A Systematic review. *World Neurosurgery, 109*, 476-486. doi:10.1016/j.wneu.2017.09.149

Shorey, S., Ang, E., Yap, J., Ng, E. D., Lau, S. T., & Chui, C. K. (2019). A Virtual counseling application using artificial intelligence for communication skills training in nursing education: Development study. *Journal of Medical Internet Research, 21*(10), e14658. doi:10.2196/14658

Simmons, A. B., & Chappell, S. G. (1988). Artificial intelligence-definition and practice. *IEEE Journal of Oceanic Engineering, 13*(2), 14-42.

Strudwick, G. (2015). Predicting nurses' use of healthcare technology using the technology acceptance model: an integrative review. *CIN: Computers, Informatics, Nursing, 33*(5), 189-198.

Teo, T. (2019). Students and teachers' intention to use technology: Assessing their measurement equivalence and structural invariance. *Journal of Educational Computing Research, 57*(1), 201-225.

Teo, T., & Zhou, M. (2014). Explaining the intention to use technology among university students: A Structural equation modeling approach. *Journal of Computing in Higher education, 26*(2), 124-142.

Teo, T., Milutinović, V., & Zhou, M. (2016). Modelling Serbian pre-service teachers' attitudes towards computer use: A SEM and MIMIC approach. *Computers & Education*, *94*, 77-88.

Thompson, W. R., Reinisch, A. J., Unterberger, M. J., & Schriefl, A. J. (2019). Artificial intelligence-assisted auscultation of heart murmurs: Validation by virtual clinical trial. *Pediatric Cardiology, 40*(3), 623-629.

Ursavaş, Ö. F., Yalçın, Y., & Bakır, E. (2019). The Effect of subjective norms on preservice and in-service teachers' behavioural intentions to use technology: A Multigroup multimodel study. *British Journal of Educational Technology, 50*(5), 2501-2519.

Venkatesh, V., & Davis, F. D. (2000). A Theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science, 46*(2), 186-204.

Wang, W. T., & Wang, C. C. (2009). An Empirical study of instructor adoption of web-based learning systems. *Computers & Education, 53*(3), 761-774.

Williams, A. M., Liu, Y., Regner, K. R., Jotterand, F., Liu, P., & Liang, M. (2018). Artificial intelligence, physiological genomics, and precision medicine. *Physiological Genomics, 50*(4), 237-243.

Winfield, A. (2019). Ethical standards in Robotics and AI. *Nature Electronics, 2*(2), 46-48.

Wu, L., Li, J. Y., & Fu, C. Y. (2011). The adoption of mobile healthcare by hospital's professionals: An Integrative perspective. *Decision Support Systems, 51*(3), 587-596.

Yang, T. C., Hwang, G. J., & Yang, S. J. H. (2013). Development of an adaptive learning system with multiple perspectives based on students' learning styles and cognitive styles. *Educational Technology & Society, 16*(4), 185-200.

Yang, Z., & Lin, M. (2019). Intelligent fan system based on big data and artificial intelligence. In *Proceedings of the 6th International Conference on Systems and Informatics (ICSAI)* (pp. 637-642). doi:10.1109/ICSAI48974.2019.9010072

Yetisen, A. K., Martinez-Hurtado, J. L., Ünal, B., Khademhosseini, A., & Butt, H. (2018). Wearables in medicine. *Advanced Materials, 30*(33), 1706910. doi:10.1002/adma.201706910

Yu, P., Li, H., & Gagnon, M. P. (2009). Health IT acceptance factors in long-term care facilities: A Cross-sectional survey. *International Journal of Medical Informatics, 78*(4), 219-229.

Zhang, H., Cocosila, M., & Archer, N. (2010). Factors of adoption of mobile information technology by homecare nurses: A Technology acceptance model 2 approach. *CIN: Computers, Informatics, Nursing, 28*(1), 49-56.

Zhao, Y., Ni, Q., & Zhou, R. (2018). What factors influence the mobile health service adoption? A Meta-analysis and the moderating role of age. *International Journal of Information Management, 43*, 342-350.

Zou, J., & Schiebinger, L. (2018). AI can be sexist and racist- It's time to make it fair. *Nature, 559,* 324-326.

Appendix A

Questionnaire item

Intention	
PU01	I believe that learning to use the AI-technology tools can better assist healthcare work.
PU02	Using the AI-technology tools would increase my healthcare work productivity.
PU03	I believe that using the AI-technology tools would enhance my professional development.
PEU01	Learning to use the AI-technology tools for healthcare is easy for me.
PEU02	My interaction with the AI tools for healthcare is clear and understandable.
PEU03	Learning to operate the AI-technology tools in the healthcare field would be easy for me.
PEU04	Using the AI-technology tools would enhance the effectiveness of my healthcare work.
SN01	My supervisor or organization believes that I should employ the AI-technology tools to assist my healthcare work in the future.
SN02	I want to learn to use the AI-technology tools because my supervisor or organization requires it.
SN03	The support from my supervisors or organization in learning to use the AI-technology tools is important to me.
SN04	The opinion of my colleagues about learning to use the AI-technology tools is important to me.
ATU01	I have a generally favorable attitude toward learning to use the AI-technology tools.
ATU02	It is a good idea to learn to use the AI-technology tools for healthcare work and personal and professional development.
BI01	I intend to learn to use the AI-technology tools for my healthcare work in the future.
BI02	I intend to learn to use the AI-technology tools for my healthcare work frequently.
BI03	I intend to adapt the AI-technology tools for healthcare work and personal and professional development.

Zhao, F., Hwang, G.-J., & Yin, G. (2021). A Result Confirmation-based Learning Behavior Analysis Framework for Exploring the Hidden Reasons behind Patterns and Strategies. *Educational Technology & Society*, *24* (1), 138–151.

A Result Confirmation-based Learning Behavior Analysis Framework for Exploring the Hidden Reasons behind Patterns and Strategies

Fuzheng Zhao[1], Gwo-Jen Hwang[2] and Chengjiu Yin[1*]
[1]Kobe University, Japan // [2]National Taiwan University of Science and Technology, Taiwan //
zhaofz635@gmail.com // gjhwang.academic@gmail.com // yin@lion.kobe-u.ac.jp
[*]Corresponding author

ABSTRACT: Educational data mining and learning analytics have become a very important topic in the field of education technology. Many frameworks have been proposed for learning analytics which make it possible to identify learning behavior patterns or strategies. However, it is difficult to understand the reason why behavior patterns occur and why certain strategies are used. In other words, all of the existing frameworks lack an important step, that is, result confirmation. In this paper, we propose a Result Confirmation-based Learning Behavior Analysis (ReCoLBA) framework, which adds a result confirmation step for exploring the hidden reasons underlying the learning patterns and strategies. Using this ReCoLBA framework, a case study was conducted which analyzed e-book reading data. In the case study, we found that the students had a tendency to delete markers after adding them. Through an investigation, we found that the students did this because they could not grasp the learning emphasis. To apply this finding, we proposed a learning strategy whereby the teacher highlights the learning emphasis before students read the learning materials. An experiment was conducted to examine the effectiveness of this strategy, and we found that it could indeed help students achieve better results, reduce repetitive behaviors and save time. The framework was therefore shown to be effective.

Keywords: Learning analytics, Learning behavior pattern, Learning Analysis framework, Result confirmation

1. Introduction

In the last decade, technologies of educational data mining and learning analytics have made rapid progress (S. Baker, 2019). With the development of algorithm techniques, more scalable algorithms that can access an increasing number of computational resources have been proposed (Hashem, Yaqoob, Anuar, Mokhtar, Gani, & Khan, 2015). Moreover, there is greater availability of large amounts of fine-grained education data than before (Dietze, Siemens, Taibi & Drachsler, 2016). Despite it being easy to collect detailed learning data from multiple resources and to analyze them in order to provide educational suggestions or recommendations, the analysis results may not be sufficient to understand the critical information (Maldonado-Mahauad, Pérez-Sanagustín, Kizilcec, Morales, & Munoz-Gama, 2018).

In other words, through the use of some convenient data in a Web-based educational environment such as an e-book system, and the use of many mature analysis techniques such as sequence analysis, some learning behavior patterns can be easily found. However, it is not easy to understand the underlying reasons for behavior patterns. For example, Li, Uosaki, Ogata, Mouri, and Yin (2018) and Yin et al. (2019) found a backtrack reading behavior pattern which showed that certain students often return immediately to previous pages when they read e-textbooks; however, it is difficult to understand the reason for this particular behavior pattern. Therefore, it is very important to understand why the learning patterns or strategies behind behaviors occur, especially in terms of the complex social and cognitive processes involved.

In the philosophy of science, confirmation is an academic term related to the effect of evidence for hypotheses (Greco, Słowiński, & Szczęch, 2016). Based on that, result confirmation is viewed in this study as a step in which the reasons underlying behavior or strategies can be determined. Currently, the confirmation of the analysis results mainly depends on the judgment of the correct probability of the results obtained by the analysis algorithm. With this help, it can be inferred to what extent the correct analysis result has been obtained, in order to confirm it (Greco et al., 2016). However, this kind of confirmation emphasizes how relevant the learning behavior is to the analysis results, rather than the reasons underlying the behavior. Kennedy and Terry (2004) pointed out that cognitive components in user log data make it difficult to match a kind of behavioral pattern or strategy with a kind of behavior reason. There may be multiple reasons for the same behavior (Misanchuk & Schwier, 1992). Besides, it is noted that there has been limited analysis result confirmation in the field of learning analytics frameworks.

The main learning analytics frameworks proposed to date are: “Wisdom-Knowledge-Information-Data” (M. Baker, 2007), “Five Steps of Analytics” (Campbell & Oblinger, 2007), “Web Analytics Objectives” (Hendricks, Plantz, & Pritchard, 2008), "Collective Applications Model" (Dron & Anderson, 2009), "Processes of Learning Analytics" (Elias, 2011), and "Learning Analytics Processes" (Chatti, Dyckhoff, Schroede, & Thüs, 2012). It is evident that these learning analytics frameworks lack the indispensable confirmation step, which is a way to provide a structured process for the learning behavior analysis (Campbell & Oblinger, 2007). Based on previous research, it is easy to see that all of these frameworks usually consist of data collection, data processes, data analysis, and data application. While these frameworks have become more sophisticated, they all still lack the analysis result confirmation step. It is therefore difficult to identify the reasons behind the learners' behavior patterns and strategies. It also leads to ambiguous guidance for the application of education practice.

Based on the commonly used definition of learning analytics (Siemens & Baker, 2011), the primary function of a learning analytics framework can be summarized as understanding and optimizing learning and the corresponding contexts by the collection, measurement, analysis, and reporting of data. Notably, the functions of each component in the framework differ greatly. In terms of data analysis, it does not have the same functions as other components. For example, Yin and Hwang (2018) summarized the goals of data analysis as prediction, structure discovery, and relationship mining. With the advances in incorporating a variety of techniques, the requirements of the framework for educational practice have become increasingly precise, and have also been synchronized with the development of analytical techniques and learning theories. However, without confirming the analysis results, numerous obstacles have arisen in practice. In particular, the current research frameworks lack a crucial function of confirming the analysis results, which leads to an imbalance between practical requirements and framework functions; that is, the existing frameworks cannot meet the requirements for increasingly precise analysis results.

In this paper, we propose the Result Confirmation-based Learning Behavior Analysis (ReCoLBA) framework which includes a function to confirm the students' behavioral patterns and strategies. The ReCoLBA framework features the integration of confirming the reasons behind learning behavior into the existing frameworks. Its functional elements include data collection, data processing, data analysis, result confirmation, and result application.

This framework can also be applied to precision education. Recently, precision education has emerged as an important idea, and has been seen as having great potential for predicting which students are at risk, and providing timely interventions. On the basis of the extraction of the same philosophy between precision medicine and precision education, Lu et al. (2018) defined the objective of precision education as the improvement of diagnosis, prediction, treatment, and prevention of learning outcomes. However, relying on the existing learning analysis frameworks, we can only obtain the analysis of the recorded learning behavior, but cannot explain why this behavior occurs. Kennedy and Terry (2004) interpreted the reason for this as the lack of cognitive components in the electronic records of student activities. Through our proposed framework, we can not only determine which students are at risk, but also confirm the reasons for such behavior, so as to avoid a one-size-fits-all learning strategy.

Using the ReCoLBA framework, a case study was conducted by analyzing e-book reading data. In this ReCoLBA-based case study, we found that the students had a tendency to delete markers after adding them. Through an investigation, we found that they did this because they could not grasp the learning emphasis. To apply this finding, we proposed a learning strategy whereby the teacher highlights the learning emphasis on the e-textbooks before the students read the learning materials. At the same time two experiments were used, one to examine the usability of the framework, and the other to verify the effectiveness of the learning strategy proposed by the confirmed analysis results. Finally, The ReCoLBA-based case study illustrates that the framework is effective, and we found that the learning strategy could help students achieve better results and save time.

2. Literature review

2.1. The learning analytics framework

The term Educational Data Mining first appeared in a workshop in 2005, and then in 2008 the First International Conference on Educational Data Mining was held (Baker & Inventado, 2014). As a sister community to educational data mining, research on learning analytics and its frameworks followed. In 2007 M. Baker presented a framework, "wisdom-knowledge-information-data," calling it a "Knowledge Continuum." This framework emphasizes data processing in which knowledge is converted into a meaningful form (M. Baker, 2007). Compared with the above abstract learning analytic framework, Campbell, DeBlois, and Oblinger (2007) proposed the "Five Steps of Analytics": capture, report, predict, act and refine, for a more simplified, practice-oriented structure procedure in the same year.

After the basic analytical framework had been constructed, coinciding with the rise of network analytic research, Hendricks et al. (2008) shifted the emphasis of the learning analytics framework study to the objectives of web analytics. From their perspective, the four operations of defining goals, measuring outcomes, using the resulting data, and sharing data for other stakeholders were identified as four objectives when using web analytics in education. These four objectives commonly constitute a learning analytics framework centered on the analysis objective.

With the recognition of the important role of social software in E-learning, Dron and Anderson (2009) noted that one kind of distinct dynamics has emerged in educational settings. It follows that group, network, and collective concepts gradually gained attention. Then, in 2009, they proposed a "Collective Applications Model" framework comprising five layers: selecting, capturing, aggregating, processing, and displaying. This framework was also classified into three cyclical phases: information gathering, information processing, and information presentation (Dron & Anderson, 2009). The cyclic structure with the head and tail connected was first put forward in this framework. In 2011, the definition of learning analytics was established at the 1st International Conference on Learning Analytics and Knowledge, followed by the learning analytics framework development in the theoretical dimension. Meanwhile, Elias provided a comprehensive learning analytics framework by summarizing the existing frameworks consisting of select, capture, aggregate & report, predict, use, refine, and share (Elias, 2011).

Despite the fact that these learning analytics frameworks were constructed in relation to the analysis process, there were still questions regarding combining them with the educational theory (Romero & Ventura, 2013). Chatti et al. (2012) introduced a successful learning theory, namely "Kolb's Experiential Learning Cycle," into the learning analytics for analysis framework. They also described a reference model for learning analytics, in light of an iterative cycle proposed as an older learning theory (Clow, 2012). This framework starts with data collection and processing, then goes into analytics and action, importantly, not ending with post-processing but entering a new cycle in a way that affects the next data collection. Similarly, in 2013, Siemens proposed a learning analytics cycle framework adopting a systems approach that includes seven components: collection, storage, data cleansing, data integration, analysis, visual presentation, and action (Siemens, 2013).

As a subfield of learning analytics, the learning behavior analysis framework basically follows the basic ideas of existing learning analytics frameworks. Although there are differences between the learning analysis frameworks mentioned above, a common feature is that each framework lacks an analysis result confirmation step. Therefore, we present the ReCoLBA framework, in which the result confirmation was added. Obviously, the distinctive function of this framework is to confirm why the behavior takes place and the strategies adopted.

2.2. Result confirmation

Confirmation is a term in the philosophy of science which is defined as the effect of evidence for a hypothesis. Here we define result confirmation as the analysis of the cause of the occurrence of behavior or strategies, which are obtained by learning analytics. The purposes of result confirmation are twofold: the first is the analysis algorithm confirmation, which is based on the probabilistic theory to infer the accuracy of the analysis results, whose evaluation metrics include support, confidence, correlation, and lift (Greco et al., 2016). With this help, it can be inferred to what extent the correct analysis result has been obtained, to complete the confirmation of the analysis result; the second is the analysis results confirmation. Generally, it is carried out using a pragmatic, mixed-methods approach (Phillips, 2006), and the analytical results are confirmed through self-reports, questionnaires, interviews, cases, and observations.

The result confirmation was designed to provide an explanation of the reasons behind the results. For example, the video click rate or duration is collected to predict future academic performance and dropout rates. If the analysis results are applied directly without confirmation, the positive effect of the results on practice is not guaranteed. Especially, solely relying on the click rate or duration is an analysis indicator, which is not causally related to academic performance, but only correlated. It has been proven that only using learning engagement to predict academic performance fails to accurately identify which student is at risk of dropping out. In fact, students who are usually considered as potentially giving up, are not necessarily those who drop out. Low engagement level has little to do with learning itself, and may be related to bad time management (Gourlay, 2017). Therefore, the real reasons behind these behavior results cannot be accurately discovered without confirmation, thus giving rise to the fuzzy specific guidance for stakeholders in education institutions.

2.3. Precision education

In the light of former US President Obama's 2015 State of the Union address where the precision medicine initiative was mentioned (Collins & Varmus, 2015; White House, Office of the Press Secretary, 2015), Hart widened the defining scope of precision medicine, including learning disabilities (LD) in the category, rather than it being limited to only biomedical diseases. Moreover, Hart (2016) argues that precision education would help researchers and practitioners understand the complex mechanisms underlying LD. Cook, Kilgus, and Burns (2018) defined precision education based on the best available evidence, and defined it as an approach to research and practice that is concerned with tailoring prevention and intervention practices to individuals. More recently, an investigation on how big data and artificial intelligence can be used to help universities more precisely understand student backgrounds was carried out (Tsai, Chen, Shiao, Ciou, & Wu, 2020). According to the results, the studies with the notion and science of precision education may enable universities to provide interventions to students for course selection and competence growth (Hart, 2016).

3. Result confirmation-based learning behavior analysis framework

As shown in Figure 1, the ReCoLBA framework, which consists of five main steps: data collection, data processing, data analysis, result confirmation, and result application, has a cyclic and iterative structure. Based on the two frameworks presented by Chatti et al. (2012) and Siemens (2013) and Kolb's Experiential Learning Cycle learning theory, we propose a Result Confirmation-based Learning Behavior Analysis (ReCoLBA) framework in this study.

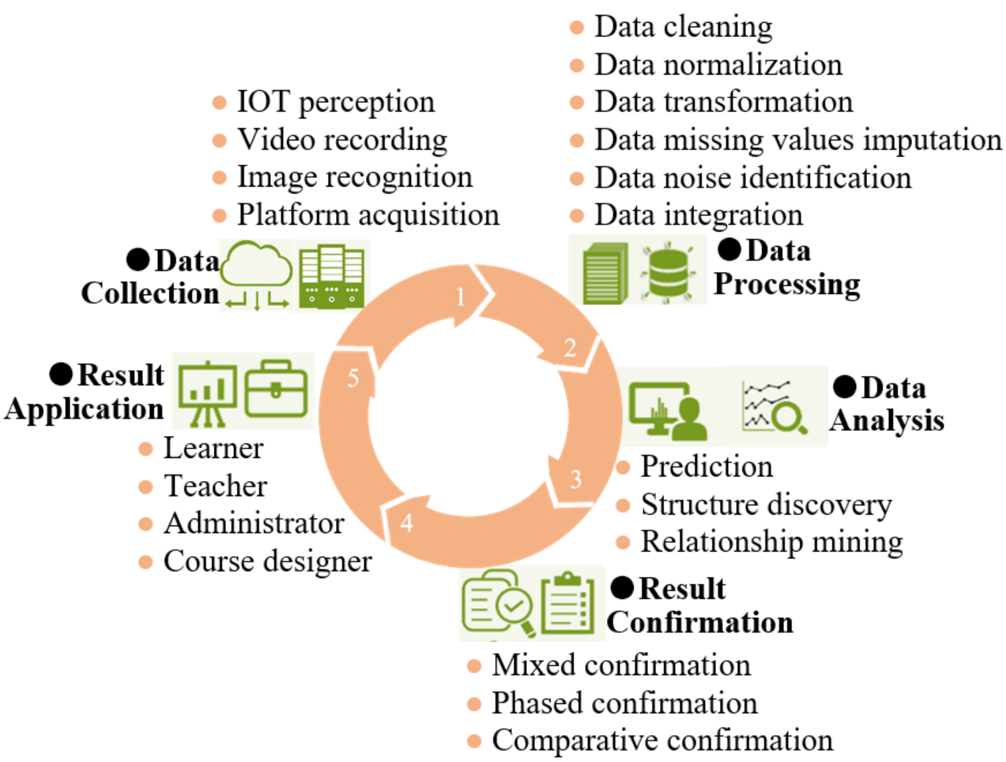


Figure 1. The ReCoLBA framework

By comparing and combining the aforementioned frameworks adopted by the ReCoLBA, while clarifying the differences and similarities between them, we highlight the innovative functions of this framework in Table 1. The most important functions of identifying and confirming the reasons underlying the learning behavior are obvious. Moreover, an important education theory, Kolb's Experiential Learning Theory, was introduced into the ReCoLBA framework. The integration of that theory provides a cyclic and iterative structure for this framework. Subsequently, a specific introduction of each step will be provided.

Table 1. The differences and similarities between the reference frameworks and the ReCoLBA framework

Name	ReCoLBA framework	Learning Analytics framework (Chatti et al., 2012)	Processes of Learning Analytics framework (Siemens, 2013)
Steps	Data collection	Data collection	Collection Storage
	Data processing	Pre-processing	Data cleaning Data Integration
	Data analysis	Analytics	Representation and visualization
	Result confirmation	None	None
	Result application	Action and Post-processing	Action

3.1. Data collection

The first primary function of this framework is to collect data from different education environments, with the main data collecting methods, such as IOT perception, video recording, image recognition, and platform acquisition. According to the report "SEDCAR, Standards for Education Data Collection and Reporting" (National Center for Education Statistics, 1991), data in educational settings or systems can be collected and analyzed by means of record extraction, surveys (mail, telephone, face-to-face), observations, experiments, and secondary data analysis. Usually, the collection methods can be divided into two categories: designed experiments and the observational approach. The former means that collectors control the data generation conditions. The latter means that collectors do not participate in the data generation process (Kantardzic, 2011). At present, data collection technologies include the following four main technical categories (Wong, 2017): Internet-of-Things (wearable devices), video recording technology (video broadcasting), learning platform acquisition technology (log data), and image recognition technology (eye-tracking).

3.2. Data processing

The second data processing step includes data cleaning, data normalization, data transformation, data missing values imputation, data noise identification, and data integration (Romero & Ventura, 2010a; García, Ramírez-Gallego, Luengo, Benítez, & Herrera, 2016). Data processing is a step that transforms raw data into a useful and efficient format (Chakrabarty, Mannan, & Cagin, 2015). The main tasks of data processing are to retrieve inaccurate records in the data set, identify incorrect or irrelevant records in the data set, and manipulate the collected data by deleting, modifying, and replacing (Wu, Zhu, Wu, & Ding, 2013).

3.3. Data analysis

The following step is data analysis, which consists of the three analysis goals of prediction, structure discovery, and relationship mining. Data analysis is an operation which is guided by certain research purposes, such as prediction, structure discovery, and relationship mining (Yin & Hwang, 2018). Each particular educational problem has its own specific objectives, so the existing analytical methods and techniques cannot be directly applied to the analysis of such data (Romero & Ventura, 2013). In other words, the analysis methods are not categorized into special research areas in the education field. The data analysis methods are divided into 11 methods according to three categories: prediction, structure discovery, and relationship mining (Yin & Hwang, 2018), as shown in Table 2.

Table 2. Goals and methods of data analysis

Goals	Prediction	Structure discovery	Relationship mining
Methods	Classification	Clustering	Association rule mining
	Regression	Factor analysis	Correlation mining
	Latent knowledge	Knowledge inference	Sequential pattern mining
	Estimation	Network Analysis	Causal data mining

3.4. Result confirmation

The obvious function point of the framework which differs from any others is the result confirmation, which is made up of mixed confirmation, phased confirmation, and comparative confirmation. This study reviewed previous findings on the learning analytics frameworks, and found that they were lacking a results confirmation step. To address the problems of confirming the analysis results, four representative research papers were identified related to how to conduct a study to confirm the analysis results. For example, to evaluate the online materials in the context of the unit of study, Phillips, Baudains, and Van (2002) confirmed existing results by using learning logs recorded in the WebCT site for students' approaches. In addition to quantitative approaches to confirming learning outcomes using statistical methods, such as comparison of the previous research in early years, an investigation was employed to confirm the situation about completing laboratories, ongoing study, and surface learning. Only relying on the understanding of the fact that the surface analysis using data-based learning analytics is insufficient to identify student learning behavior, a case study of how qualitative data provide rich information to confirm the analysis results was carried out by Phillips (2006), and a comparative confirmation using the data generated from the study subjects in different periods was used to discover the changing laws (Kennedy & Terry, 2004; Li et al., 2018). In response to the insufficient interpretation of data analysis for the learning behavior, they interviewed the learners to confirm the potential interactions after the data analysis. By carrying out investigations, this approach with phased confirmation steps can identify the potential reasons underlying the existing analysis, such as why students repeat the learning behavior.

Based on the specific confirmation methods presented in these four papers, this study summarizes them into the three categories of mixed confirmation, comparative confirmation, and phased confirmation.

- Mixed confirmation. This is a method of confirming the analysis results through a combination of quantitative and qualitative studies. It not only includes using evaluation metrics of the analysis algorithm, but also involves some qualitative methods such as documentation, staff interviews, observation of students in laboratory classes, and a student survey (Phillips et al., 2002).
- Comparative confirmation. There are two confirmation dimensions of the comparative confirmation method. The first is the time dimension. The comparison of the similarities and differences is used to discover the changing laws, by means of collecting the data generated from the study subjects in different periods. Data collection intervals are chosen, based on semesters or school years. The second is the spatial dimension. For example, students' learning behavior data from different schools are comparatively analyzed for results confirmation (Phillips, 2006).
- Phased confirmation. The results confirmation involves two analytic stages. The statistical analysis is usually adopted in the first stage. If the variables of interest had occurred, or the analysis results had abnormal values, then they would move to the second stage. The focuses on interesting variables or abnormal values will be analyzed again. The confirming method includes interviews and questionnaires (Kennedy & Terry, 2004; Li et al., 2018).

3.5. Result application

The last one is the result application step which involves four stakeholders: learner, teacher, administrator and course designer. It is noteworthy that the final goal of the ReCoLBA framework is to apply the analysis results to educational practice. Different roles can derive different benefits from the confirmed results of learning analytics. For example, the confirmed analysis results could help students share their learning experience, help teachers master students' learning behavior and provide them with timely support, help manager administrators to evaluate teachers and students, and help course designers to improve the course content and instructional materials according to the confirmed analysis results (Yin & Hwang, 2018).

4. Experimental design

Using the ReCoLBA framework, we designed a case study including two experiments, one aiming to examine the usability of the framework, and the other focusing on verifying the effectiveness of a learning strategy proposed by the confirmed analysis results.

4.1. Experiment for verifying the usability of the ReCoLBA framework

To examine the usability of the proposed framework, a ReCoLBA-based case study was conducted. To this end, 48 participants were recruited and assigned to read academic papers using an e-book system. There are five parts in this experiment, namely using the e-book system to collect students' reading behavior, performing five steps of data processing, adopting behavioral sequential analysis for data analysis, utilizing a mixed confirmation method to confirm the analysis results, and applying a learning strategy by an intervention.

4.1.1. Data collection in the case study

An e-book system was developed corresponding with the functions of the ReCoLBA framework, to evaluate the framework usability. As shown in Figure 2, this system can collect the students' reading behavior when reading learning materials, such as (1) page-turning, (2) zoom+/-, (3) writing a memo, (4) adding or deleting an underline, and (5) adding or deleting a highlight. Most importantly, all reading behavior actions were recorded in the form of reading action logs. Moreover, the e-book system can provide automatically coded action logs as one of the convenient functions to reduce subsequent data processing. In this way, the usual manual coding process completed by programmers can be avoided.

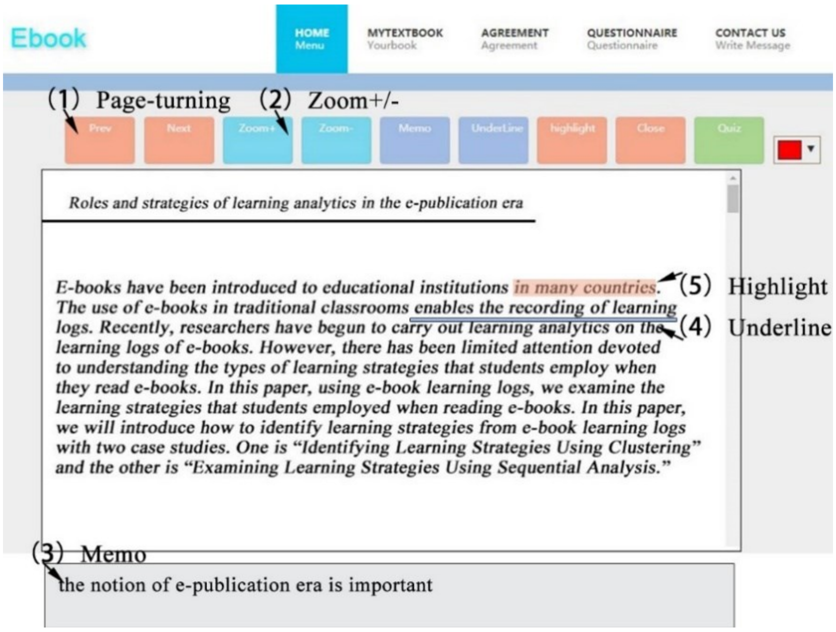


Figure 2. The E-book system interface

A total of 4,748 records were collected from 60 graduate students, who were asked to complete reading academic papers in the e-book system within 90 mins in this experiment. Their reading behavior data, which consisted of user ID, operation name, page number, and action time, were stored in the database for analyzing their learning behavior patterns or strategies, as shown in Table 3.

Table 3. Sample action log

User ID	Operation name	Page number	Action time
Student 1	NEXT page:0	19	2020/7/10 12:05:18
Student 2	PREV page:8	16	2020/7/10 12:07:31
Student 3	ADD HL page:4	16	2020/7/10 12:36:54

4.1.2. Data processing in the case study

This study summarizes data processing as the following four points. The first is data transformation. Through the statistical processing, the sum of the students' reading behavior including adding or deleting underlines, adding or deleting highlights, and adding or deleting memos, was respectively counted. The second point is data cleaning. If the following learning phenomenon or behavior occurred during the learning activity, it was viewed as an invalid record. For example, if the longest duration between two learning behaviors exceeded 20 minutes, then the record was invalid as it indicated that no reading activities had occurred because the student did not conduct any learning behavior within 20 minutes. Besides, incomplete records were filtered. The third point is missing values imputation and noise identification. The valid sample size changed from 60 to 47 by identifying and discarding missing values data and noisy data. The fourth point is data normalization. There were some data completed by those who had an invalid preview, for example, some students who completed the preview of the lesson (read the learning content before class) less than 3 minutes before the class. In that case, the preview of the lesson was viewed as invalid. The final point is data integration. The objective of data processing is mainly from the e-book system, so that the sole data generating source avoids the need to integrate data from different data sources, which makes it unnecessary to adopt data transformation and integration in the data processing.

4.1.3. Data analysis in the case study

In the case study, behavioral sequential analysis was adopted to gain a detailed understanding of progressive learning behavioral patterns (Bakeman & Gottman, 1997; Hou, 2012). The progressive learning behavior patterns were obtained by sequential analysis, as shown in Figure 3. The squares represent learning behavior, and the arrow-lines and numbers represent the direction and extent to which the behavior is associated with other behaviors. Rounded curves represent the association of the behavior with itself. The result is significant at the < .05 level when the *z*-value is greater than 1.96 (Bakeman & Gottman, 1997; Hou, 2012).

In Figure 3, it is obvious that "HIGHLIGHT" is mutually associated with "DEL HIGHLIGHT" and "DEL UNDERLINE," and "UNDERLINE" is mutually associated with "DEL HIGHLIGHT" and "UNDERLINE." Meanwhile, "HIGHLIGHT," "DEL HIGHLIGHT," "DEL UNDERLINE," and "UNDERLINE" are respectively associated with themselves. "BOOKMARKER" is mutually associated with "DEL BOOKMARKER." Finally, "BOOKMARKER" has a one-way association with "UNDERLINE" and "MEMO." The association between this repetitive behavior was confirmed to be significant, using the Z-score binomial test parameters in the sequential analysis method. However, this is not sufficient to understand why these patterns and strategies occurred, and especially to understand why some students exhibit some repetitive behavior of the same operation. For example, some students deleted UNDERLINE after adding an UNDERLINE, and some deleted the HIGHLIGHT after adding a HIGHLIGHT.

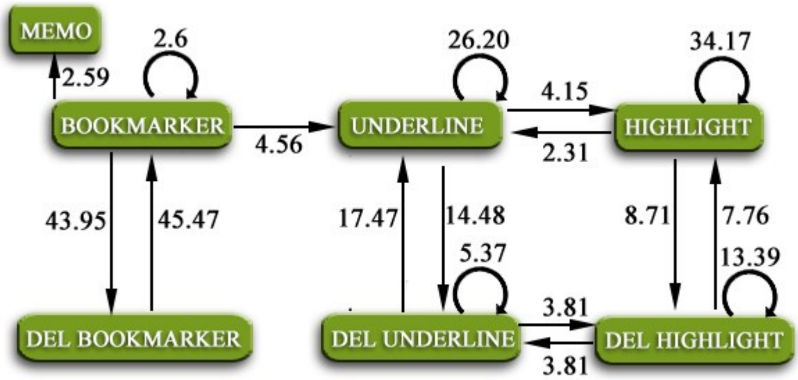


Figure 3. The students' progressive learning behavioral patterns

4.1.4. Resulting confirmation in the case study

We identified the repetitive learning behavior patterns as "Deleting a HIGHLIGHT after adding the HIGHLIGHT," "After adding an UNDERLINE, delete the UNDERLINE." However, it was difficult to understand the reason why these behavior patterns happened. Therefore, it was necessary to conduct the result confirmation step. To this end, an investigation system was used to investigate the participants who had repetitive learning behavior patterns, as shown in Figure 4. The participants were investigated to answer the questions that matched their behavior such as, "Why did you adopt "After adding a HIGHLIGHT, delete the HIGHLIGHT," and why did you adopt "After adding an UNDERLINE, delete the UNDERLINE." Samples of extracts from their answers are shown in Table 4.

In this investigation, only if common results were gained by two researchers was the coding of the students' answers accepted. The notion for the data analysis of up to 24 participants was extracted, and the reference basis is strong. After the investigation, it was confirmed that it was difficult for some students to identify which were the keywords while reading. Other students often read repeatedly to understand the main idea of the papers because they had difficulty finding the most important content. It was concluded that the repetitive operation of markers happened because they could not correctly identify the learning emphasis. Besides, the marking method, such as underline and highlight, is beneficial for students to mark the important content, and these choices are mainly related to personal interest.

Figure 4. The investigation system interface

Table 4. Samples of participants' answer content

Participant	Answer
Student 1	When I first read it, I highlighted the content that I thought was important. However, going back to read it a second time, I realized that it was not the most important content anymore, so I decided to delete it.
Student 2	It is useful to deepen my understanding of the paper.
Student 3	When I read the sentences, from my perspective, it is more suitable for the question.
Student 4	I subsequently added an underline after the highlight, which is my way of distinguishing the emphasis.
Student 5	That is a mistake in the highlight position owing to not being proficient in operating the platform.

4.1.5. Resulting application in the case study

Through the ReColBA-based case study, it was found that the reason why students exhibit the repetitive behavior of the same operation is that they could not find the learning emphasis. The e-textbooks lack any marking of the learning emphasis, which is not conducive to students' understanding. Therefore, we proposed a learning strategy that marked the learning emphasis in the e-textbooks using underlines and highlights, as shown in Figure 5.

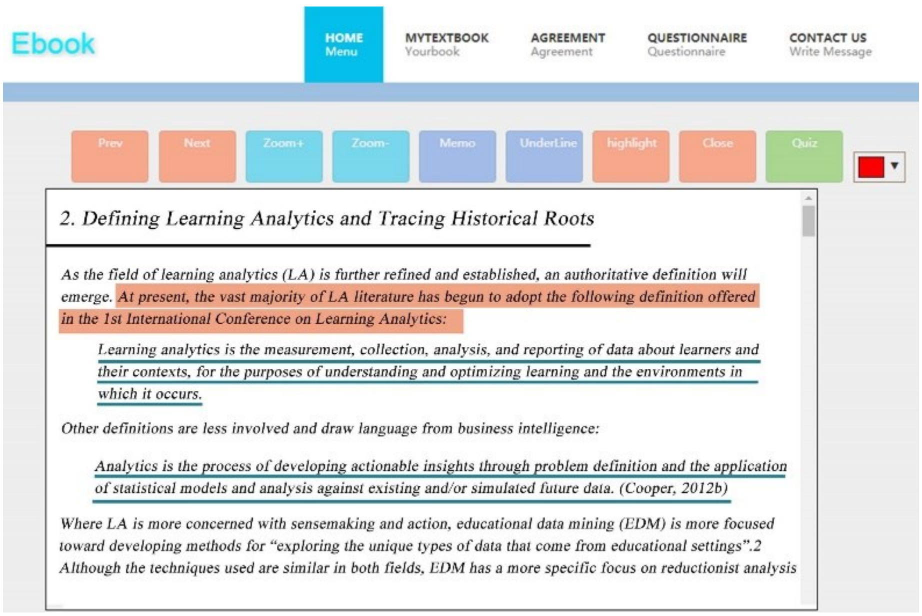


Figure 5. The e-textbook with marking of the learning emphasis

4.2. Experiment for verifying the learning effectiveness of the ReCoLBA framework

To verify the learning effectiveness of the learning strategy discovered by the framework, we conducted a verification experiment. The specific steps are as follows.

4.2.1. Participants

A total of 80 graduate students from a university's graduate school were recruited to participate in this experiment. As the experimental group, 43 students read an e-textbook via the e-book system, using the marking learning emphasis method. Another 37 students in the control group adopted the same experiment conditions, except that they did not use the marking learning emphasis technique.

4.2.2. Measuring methods

The measuring methods consisted of a pre-test, a post-test, and an interview. Besides, the frequency and duration of learning behavior which were automatically recorded by the e-book system were used to evaluate the effectiveness of the learning strategy. Firstly, the pre-test was to evaluate whether both groups had equivalent prior knowledge of the upcoming learning content. It included 10 multiple-choice items. Secondly, the post-test was to measure whether the marking learning emphasis method was helpful for students' learning achievement. It also included 10 multiple-choice items, which were related to the core learning emphasis of the article. The total score was 100 in both the pre-test and post-test. Thirdly, a 30-minute semi-structured interview was conducted for the experimental group which was recorded verbatim. The following questions were asked: What do you think of this learning method? What is the difference between this way of marking the key points of the paper and your previous learning method? Two researchers were invited to analyze the interview data, and to determine the participants' attitudes towards the proposed learning strategies through the extracted core keywords from the interview content.

4.2.3. Experimental procedure

The learning emphasis aimed to help students understand the definition, historical development, and application fields of Learning Analytics, which is a unit of an Education Technology course in the university.

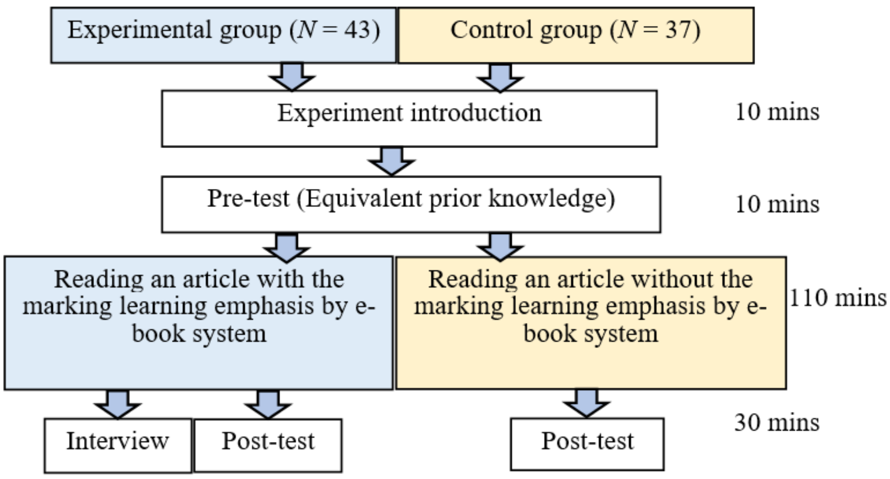


Figure 6. Experiment design diagram

As shown in Figure 6, before the experiment, the students were given a pre-test to test their prior knowledge of the learning emphasis. Subsequently, we introduced the learning emphasis and e-book system, then conducted a 110-minute learning activity. Both the experimental group and the control group were required to use the e-book system functions at will, such as page-turning, underlining, highlighting, and making memos. Unlike the control group, the e-textbook for the experiment group was already underlined and highlighted, to indicate the learning emphasis. After the learning activity, a post-test was conducted, and the experimental group students were asked to take part in an interview.

5. Experiment results

First, we analyzed the participants' learning achievement. The effectiveness of the proposed method was examined to ensure whether this approach could improve the students' learning performance. Based on the pre-test, the standard deviations and mean values were 20.471 and 60.00 for the control group, and 15.596 and 78.92 for the experimental group. According to the *t*-test result ($t = 2.897$, $p > .05$), as shown in Table 5, there was no significant difference between the two groups. It was therefore found that they had equivalent prior knowledge before the learning activity.

Table 5. Descriptive data and *t*-test result of the pre-test results

Variable	Group	*N*	Mean	*SD*	*t*
Pre-test	Experimental group	43	60.00	20.471	2.897
	Control group	37	78.92	15.596	

The post-test was taken after the learning activity. We used the one-way analysis of covariance (ANCOVA) to evaluate the learning achievement of the two groups, setting the groups as a fixed factor, the pre-test scores as the covariate, and the post-test scores as the dependent variable. The data met the ANCOVA requirements. The Levene's test of equality of error variances ($F = 1.409$, $p > .05$) indicated that the assumption of regression homogeneity was followed. The results of the ANCOVA ($F = 18.424$, $p < .01$) in Table 6 determined that the experimental group achieved better results. It was concluded that the marking learning emphasis method was beneficial for the students to understand the learning emphasis.

Table 6. Descriptive data and one-way ANCOVA result of the post-test results

Variable	Group	*N*	Mean	*SD*	Adjusted mean	*F*
Post-test	Experimental group	43	88.60	13.378	89.74	18.424*
	Control group	37	77.62	11.526	76.29	

Note. $^{*}p < .01$.

Second, we analyzed the system reading time. To better understand the effectiveness of this method, the ANCOVA was also used to analyze the reading time of the two groups. The Levene's test of equality of error variances ($F = 1.409$, $p > .05$) was not violated. From the ANCOVA results in Table 7, it can be seen that the means and standard deviations of the experimental group are 0:41:06 and 0:21:21, and for the control group they are 1:12:27 and 0:26:26. There was a significant difference ($F = 41.731$, $p < .01$) between the two groups. It was clear that the experimental learners spent less system reading time finishing the learning task than those who used the conventional method. It was found that it was helpful to save time compared with the students who did not use the marking learning emphasis method.

Table 7. Descriptive data and one-way ANCOVA result of reading time

Variable	Group	*N*	Mean	*SD*	Adjusted mean	*F*
Reading time	Experimental group	43	0:41:06	0:21:21	0:38:12	41.731*
	Control group	37	1:12:27	0:26:26	1:16:32	

Note. $^{*}p < .01$.

Third, we analyzed the results of the Highlight and Underline operations. To examine the effectiveness of the intervention by marking the learning emphasis, the deleting rate of highlight and underline operations was viewed as a dependent variable, to test whether there were significant changes in the repetitive behaviors. In terms of the deleting rate of the highlight results, it was found that the standard deviation and mean values were 0.214 and 0.100 for the control group, and 0.040 and 0.010 for the experimental group. The *t*-test result ($t = 0.016$, $p < .05$), as shown in Table 8, indicated a significant difference between the two groups. For the underline results, it can be seen that the mean values and standard deviations of the experimental group are 0.011and 0.053, and for the control group they are 0.071 and 0.158. There was a significant difference ($t = 0.032$, $p < .05$) between the two groups. It is concluded that a significantly decreasing tendency of the deleting rate occurs, which illustrates that the repetitive behaviors were reduced.

Table 8. Descriptive data and *t*-test result of the deleting rate of the Highlight and Underline results

Variable	Group	*N*	Mean	*SD*	*t*
Highlight	Experimental group	43	0.010	0.040	0.016
	Control group	37	0.100	0.214	
Underline	Experimental group	43	0.011	0.053	0.032
	Control group	37	0.071	0.158	

Fourth, we analyzed the interview data. After the experiment, an interview was conducted on the topic of how to evaluate the learning methods, and 43 interview transcripts were obtained. It was concluded that 42 participants gave positive evaluations and one participant had a neutral attitude. Two dimensions constituted a positive evaluation, namely Knowledge comprehensibility (27 participants) and Reading convenience (32 participants). In all, 27 participants used the key items "Good for memorizing content" (14 participants), "Strong visual guidance" (4 participants), "Helps me understand the content" (6 participants) and "Convenient for future work" (3 participants) to describe what benefits it brought. In terms of reading convenience, 32 participants shared their positive evaluations. Four participants mentioned that this method could increase reading convenience, 18 argued that they could read better with the help of the reading emphasis marked by underlining or highlighting, and 10 maintained that the efficiency gains and time-saving were the most important benefits of the method.

6. Conclusions

The existing research clearly indicates that the analysis results have the hidden probability of creating misunderstandings of students' behavior in practice, especially in the case of lacking result confirmation. Owing to the cognitive element and multiple interpretations of learning behavior, result confirmation requires not only providing the patterns or strategies for learning behavior, but also identifying the reasons behind their occurrence, and only by doing so will the analysis results be better applied in practice.

A case study was then conducted to verify the usability of the ReCoLBA framework. Based on the findings concluded from the application experiment, the ReCoLBA framework has been successfully verified. It can help identify why certain learning behaviors occur and the strategies adopted, and avoids applying analysis results without confirmation. The investigation in the result confirmation step was different from that usually employed in the domain of human science. This investigation method is combined with the analysis technique, and the subjects that needed to be studied were based on the primary analysis results, which was deduced from the data.

The confirming function designed in the ReCoLBA framework enables researchers to have an opportunity to access the reasons underlying the learning behavior. Through the application of the framework, it was found that the reason for the students repeatedly adding and deleting underlines or highlights was the lack of learning emphasis. Based on the findings, an e-textbook with the learning emphasis marked with underlines and highlights was designed and developed. To verify the effectiveness of the revised learning method of previously giving the learning emphasis, an experiment was conducted. From the results of the *t*-test, one-way ANCOVA, and interviews, it was obvious that the marking learning emphasis method could help students improve their learning achievement and save time when using the revised e-textbook. Through verification, we found that our framework can help teachers discover which students are at risk of failing the course, and effectively confirm the results of the behavioral analysis. Specific interference strategies are then proposed for confirmed learning behavior, which is consistent with the goal of precision education.

Through a literature review, it was found that there are three kinds of methods to confirm the analysis results, that is, mixed confirmation, phased confirmation, and comparative confirmation. We subsequently integrated these three methods, which can be used separately for confirming analysis results, as a new step in the ReCoLBA framework. In that case, there is more space left to discuss the other confirmation method. It is noted that the ReCoLBA framework can help teachers to offer at-risk students a precise intervention by using a precise guiding strategy that is not only limited to teachers, but is also suitable for the different stakeholders in education, such as the precise management by administrators, precise self-regulation by learners, and precise course design by course designers.

Acknowledgement

Part of this research work was supported by the Grant-in-Aid for Scientific Research No.18H04125 and No. 16H06304 from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) in Japan.

References

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An Introduction to sequential analysis*. Cambridge, UK: Cambridge university press.

Baker, B. M. (2007). *A Conceptual framework for making knowledge actionable through capital formation* (Unpublished doctoral dissertation). University of Maryland University College.

Baker, R. S. (2019). Challenges for the future of educational data mining: The Baker learning analytics prizes. *JEDM Journal of Educational Data Mining*, *11*(1), 1-17.

Baker, R. S., & Inventado, P. S. (2014). Chapter 4: Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning Analytics: From Research to Practice* (pp. 61-75). New York, NY: The Springer Science+Business Media Company.

Campbell, J. P., & Oblinger, D. G. (2007). Academic analytics. *EDUCAUSE white paper*. Retrieved from https://net.educause.edu/ir/library/pdf/PUB6101.pdf

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics: A New tool for a new era. *EDUCAUSE Review*, *42*(4), 40-57.

Chakrabarty, A., Mannan, S., & Cagin, T. (2015). *Multiscale modeling for process safety applications*. Waltham, MA: The Butterworth-Heinemann Company.

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A Reference model for learning analytics. *International Journal of Technology Enhanced Learning*, *4*(5), 318-331.

Clow, D. (2012). The Learning analytics cycle: closing the loop effectively. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 134-138). New York, NY: Association for Computing Machinery.

Collins, F. S., & Varmus, H. (2015). A New initiative on precision medicine. *New England Journal of Medicine, 372*, 793–795.

Conde, M. Á., & Hernández-García, Á. (2015). Learning analytics for educational decision making. *Computers in Human Behavior*, *47*(7), 1-3.

Cook, C. R., Kilgus, S. P., & Burns, M. K. (2018). Advancing the science and practice of precision education to enhance student outcomes. *Journal of School Psychology, 66*(11), 4-10.

Dietze, S., Siemens, G., Taibi, D., & Drachsler, H. (2016). Datasets for learning analytics. *Journal of Learning Analytics*, *3*(2), 307-311.

Dron, J., & Anderson, T. (2009). On the design of collective applications. In *Proceedings of the 2009 International Conference on Computational Science and Engineering* (pp. 368-374). Vancouver, BC: IEEE.

Elias, T. (2011). Learning Analytics: Definitions, Processes and Potential. Retrieved from http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf.

García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M., & Herrera, F. (2016). Big data preprocessing: Methods and prospects. *Big Data Analytics, 1*(1), 9-31.

Gourlay, L. (2017). Student engagement, "learnification" and the sociomaterial: Critical perspectives on higher education policy. *Higher Education Policy*, *30*(1), 23-34.

Greco, S., Słowiński, R., & Szczęch, I. (2016). Measures of rule interestingness in various perspectives of confirmation. *Information Sciences*, *346*(1), 216-235.

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The Rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, *47*(7), 98-115.

Hart, S. A. (2016). Precision education initiative: Moving toward personalized education. *Mind, Brain, and Education, 10*(4), 209-211.

Hendricks, M., Plantz, M. C., & Pritchard, K. J. (2008). Measuring outcomes of United Way–funded programs: Expectations and reality. *New Directions for Evaluation*, *119*(9), 13-35.

Hou, H. T. (2012). Exploring the behavioral patterns of learners in an educational massively multiple online role-playing game (MMORPG). *Computers & Education*, *58*(4), 1225-1233.

Kantardzic, M. (2011). *Data mining concepts models methods and algorithms*. Hoboken, NJ: Wiley-IEEE Press.

Kennedy, G. E., & Judd, T. S. (2004). Making sense of audit trail data. *Australasian Journal of Educational Technology*, *20*(1), 18-32.

Li, L., Uosaki, N., Ogata, H., Mouri, K., Yin, C. (2018) Analysis of Behavior Sequences of Students by Using Learning Logs of Digital Books. In *Proceedings of the 26th International Conference on Computers in Education* (pp. 367–376). Manila, Philippines: APSCE.

Lu, O. H., Huang, A. Y., Huang, J. C., Lin, A. J., Ogata, H., & Yang, S. J. (2018). Applying learning analytics for the early prediction of Students' academic performance in blended learning. *Educational Technology & Society, 21*(2), 220-232.

Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R. F., Morales, N., & Munoz-Gama, J. (2018). Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses. *Computers in Human Behavior*, *80*(3), 179-196.

Misanchuk, E. R., & Schwier, R. A. (1992). Representing interactive multimedia and hypermedia audit trails. *Journal of Educational Multimedia and Hypermedia*, *1*(3), 35-72.

National Center for Education Statistics. (1991). SEDCAR (*Standards for Education Data Collection and Reporting*). Rockville, MD: Westat, Inc.

Phillips, R. A. (2006). Tools used in Learning Management Systems: Analysis of WebCT usage logs. In L. Markauskaite, P. Goodyear & P. Reimann (Eds.), *Proceedings of the 23rd Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education: Who's Learning? Whose Technology?* (pp. 663-673). Sydney, Australia: Sydney University Press.

Phillips, R. A., Baudains, C., & Van Keulen, M. (2002, December). An Evaluation of student learning in a web-supported unit on plant diversity. In A. Williamson, C. Gunn, A. Young & T. Clear (Eds.), *Proceedings of the 19th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education* (pp. 525-534). Auckland, NZ: UNITEC Institute of Technology.

Romero, C., & Ventura, S. (2010). Educational data mining: A Review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40*(6), 601-618.

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3*(1), 12-27.

Siemens, G. (2013). Learning analytics: The Emergence of a discipline. *American Behavioral Scientist*, *57*(10), 1380-1400.

Tsai, S. C., Chen, C. H., Shiao, Y. T., Ciou, J. S., & Wu, T. N. (2020). Precision education with statistical learning and deep learning: a case study in Taiwan. *International Journal of Educational Technology in Higher Education, 17*(4), 1-13.

White House, Office of the Press Secretary. (2015). *Fact sheet: President Obama's Precision Medicine Initiative*. Retrieved from https://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative.

Wong, B. T. M. (2017). Learning analytics in higher education: An Analysis of case studies. *Asian Association of Open Universities Journal*, *12*(1), 21-40.

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2013). Data mining with big data. *IEEE transactions on knowledge and data engineering*, *26*(1), 97-107.

Yin, C., & Hwang, G. J. (2018). Roles and strategies of learning analytics in the e-publication era. *Knowledge Management & E-Learning: An International Journal*, *10*(4), 455-468.

Yin, C., Yamada, M., Oi, M., Shimada, A., Okubo, F., Kojima, K., & Ogata, H. (2019). Exploring the relationships between reading behavior patterns and learning outcomes based on log data from e-books: A Human factor approach. *International Journal of Human–Computer Interaction*, *35*(4), 313-322.

Yang, C. C. Y., Chen, I. Y. L., & Ogata, H. (2021). Toward Precision Education: Educational Data Mining and Learning Analytics for Identifying Students' Learning Patterns with Ebook Systems. *Educational Technology & Society*, *24* (1), 152–163.

Toward Precision Education: Educational Data Mining and Learning Analytics for Identifying Students' Learning Patterns with Ebook Systems

Christopher C. Y. Yang[1*], Irene Y. L. Chen[2] and Hiroaki Ogata[3]

[1]Graduate School of Informatics, Kyoto University, Japan // [2]Department of Accounting, National Changhua University of Education, Taiwan // [3]Academic Center for Computing and Media Studies, Kyoto University, Japan // yang.yuan.57e@st.kyoto-u.ac.jp // irene@cc.ncue.edu.tw // hiroaki.ogata@gmail.com

[*]Corresponding author

ABSTRACT: Precision education is now recognized as a new challenge of applying artificial intelligence, machine learning, and learning analytics to improve both learning performance and teaching quality. To promote precision education, digital learning platforms have been widely used to collect educational records of students' behavior, performance, and other types of interaction. On the other hand, the increasing volume of students' learning behavioral data in virtual learning environments provides opportunities for mining data on these students' learning patterns. Accordingly, identifying students' online learning patterns on various digital learning platforms has drawn the interest of the learning analytics and educational data mining research communities. In this study, the authors applied data analytics methods to examine the learning patterns of students using an ebook system for one semester in an undergraduate course. The authors used a clustering approach to identify subgroups of students with different learning patterns. Several subgroups were identified, and the students' learning patterns in each subgroup were determined accordingly. In addition, the association between these students' learning patterns and their learning outcomes from the course was investigated. The findings of this study provide educators opportunities to predict students' learning outcomes by analyzing their online learning behaviors and providing timely intervention for improving their learning experience, which achieves one of the goals of learning analytics as part of precision education.

Keywords: Precision education, Learning analytics, Educational data mining, Learning pattern, Ebook learning log

1. Introduction

Precision education (Yang, 2019) has been recognized as a new challenge of applying artificial intelligence, machine learning, and learning analytics to improve both learning performance and teaching quality. To facilitate precision education, digital learning platforms have been reported to play an important role to collect the educational records of student's online and offline learning behaviors, performances, and other types of interactions (Maldonado-Mahauad et al., 2018). Utilizing the increasing volume of education data collected from various digital learning platforms, Siemens and Long (2011) identified two areas, educational data mining (EDM) and learning analytics (LA), which enable the integration and investigation of big data capabilities in education. Consequently, analyzing students' online learning behavior and identifying subgroups of students with certain learning patterns on various digital learning platforms have drawn considerable attention from the LA and EDM research communities.

LA has been identified as a conceptual framework for the analysis of student behavior and includes the prediction of students' learning performance, the process development of education data analysis (Hwang, Chu & Yin, 2017), data collection, and timely interventions (Hwang, 2014). One of the goals of LA as a part of precision education has been indicated to the prediction of students' learning outcomes from the course by analyzing their behaviors of online learning and providing timely intervention for improvement of their learning experience (Lu et al., 2018). LA approaches typically rely on educational data collected from users' interactions with information and communication technologies, such as Learning Management System (LMS) or social media (Gašević et al., 2016). The relevant literature has indicated that the development of education technologies that utilize education data gathered from learners has enabled various LA approaches to be used to help students succeed in various educational contexts (Wong, 2017; Kumar & Kumar, 2018). The analytics results will not only increase benefits for educators and learners but also present considerable potential for optimizing institutional processes (Colvin et al., 2015). Data mining techniques are commonly applied to identify patterns among students based on education data (Baker & Inventado, 2014). The interpretation of these identified patterns provides insight into learning and teaching processes, predicts students' learning outcomes from the course, suggests supportive interventions, and facilitates decision-making on resource allocation (Gašević et al., 2016). LA and EDM constitute a well-established framework that utilizes certain methods for successively gathering, processing, reporting, and acting on machine-readable data from learners to advance the educational environment (Papamitsiou & Economides, 2014) and demonstrate considerable potential for understanding and optimizing the learning process (Baker & Inventado, 2014).

However, without a series of data analytics methods for the identification of subgroups of students and their learning patterns using LA and EDM approaches, it has been demonstrated in previous studies that students have difficulty with accurately identifying and describing how they studied and what strategies they applied to previous learning activities (Zhou & Winne, 2012). Furthermore, the relevant literature indicates that students often fail to adjust the learning strategies they previously adopted to better address changing learning situations (Lust, Elen & Clarebout, 2013). Consequently, students tend to employ suboptimal learning strategies (Winne & Jamieson-Noel, 2003) when encountering unfamiliar learning situations.

To better understand how students learn and behave in these learning environments, in this study, the authors applied data analytic methods using the LA and EDM approaches to identify subgroups of students with various learning patterns utilizing an ebook system. The aim of the study was to explore student learning while an ebook system was used for both teaching and learning support. A clustering approach was employed to identify subgroups of students with different patterns of ebook learning behavior. Moreover, a statistical analysis was performed to investigate the associations between the identified subgroups of students and their learning outcomes from the course. Accordingly, the authors aimed to answer the following research questions:

RQ1: How many subgroups of students with different patterns of learning can be identified when they utilize an ebook system?

RQ2: What is the association between these subgroups of students and their learning outcomes from the course when learning with an ebook system?

2. Literature review

2.1. LA and EDM

Higher education has seen rapid development with the integration of the internet and various web-based technologies. With the rapid development and use of digital learning platforms, such as OpenCourseWare (OCW) and massive open online courses (MOOCs), education has experienced substantial growth in the volume of data derived from students' interactions with technology and their personal and academic profiles (Ferguson, 2012). The data collected from various learning activities have been used to develop predictive models of user behavior with increasing frequency in areas such as marketing, financial markets, sports, and health. Given the richness of the data that are collected in various educational settings, a growing number of educational institutions has been applying LA and EDM to support learners' strategic planning and decision-making (Chen et al., 2020). In recent years, many postsecondary educational institutions have started utilizing the data from learning activities they designed for their courses, combining LA approaches with data mining algorithms to better understand students' learning processes and their successes during the course (Pardo, Han & Ellis, 2016).

The uses of LA and EDM can provide potentially useful information from large volumes of unstructured data (Chen et al., 2020). The application of LA and EDM can promote and improve the design of online learning platforms, learning materials, and activities to ensure greater educational effectiveness and optimize learning environments (Greller & Drachsler, 2012). The obtained analysis results provide opportunities for course instructors and students to examine and adjust their strategies of teaching and learning, respectively. Consequently, with the constant adoption of new strategies, new education data will continuously be generated, leading to new analysis results for course instructors and students (Hwang, Chu & Yin, 2017).

The data collected from various digital learning platforms are now being utilized within various supportive environments. Student behavior modeling has received considerable attention in the field of EDM (Papamitsiou & Economides, 2014). Regarding LA and EDM approaches, the clustering and classification of education data seem to be the techniques most frequently used to measure and interpret student online learning behaviors when they are interacting with a digital learning platform. For example, Krumm et al. (2014) proposed an early warning system to detect at-risk students who might fail a course using students' learning data gathered through their interactions with various digital learning platforms. Hu, Lo and Shih (2014) developed an early warning system using a decision tree classifier. Corrin & De Barba (2015) propose a reporting system that presents visualized information to support students and instructors by reflecting on students' learning processes during a given period. Romero et al. (2013) used a sequential minimal optimization classification approach and students' learning behavioral data before a midterm exam to achieve the highest accuracy for predicting students' final learning performance.

The analysis of education data promotes the prediction of academic performance, the implementation of LA, and the identification and improvement of students' behavioral patterns and performance. These research fields focus on students and emphasize the role of contemporary technologies in improving their learning experience and performance (Chen et al., 2020). Nevertheless, most related researchers have extracted data from OCW, MOOCs, or LMS; very few have investigated students' learning behavioral data from ebook systems in particular.

2.2. Identification of students' learning patterns

Although applying digital learning platforms to support teaching has become common in institutions of higher education, it is still difficult for most stakeholders using these digital platforms in their classes to effectively and precisely follow how students actually interact with the given online learning materials, how they behave under a given learning activity, or how they adjust their learning behavioral patterns when engaging in certain learning activities. One of the emerging issues in the research fields of LA and EDM has been pointed to finding appropriate methods of extracting meaning from the education data (Chen et al., 2020). Most digital learning platforms do not automatically include the advanced tools required for applying LA or EDM approaches, and stakeholders have indicated that utilizing these tools is too complex as they have features that are well beyond the scope of what teachers require (Romero et al., 2016). Therefore, the need to employ new LA and EDM approaches and develop the corresponding tools for stakeholders to simply observe the behaviors and interaction patterns of students conducting certain online learning activities has been raised as a critical issue (Juhaňák, Zounek & Rohlíková, 2019). Recently, many studies have proposed different combinations of analytics methods that can be adopted to explore students' learning patterns using LA and EDM techniques in various educational settings.

Yang et al. (2019) applied the k-means clustering technique to explore the learning patterns of 1,326 undergraduate students, utilizing eight learning features that were extracted from the students' interactions with an ebook system. In the study, five subgroups of students are identified based on the eight online learning features. The differences between the subgroups in terms of their learning behaviors and learning outcomes were reported.

Similarly, in the context of using an ebook system for learning support, Yin et al. (2019) apply LA and EDM approaches in an undergraduate course to discover how learning patterns differ across 98 students taking online courses using an ebook system. In their study, based on the results of the k-means clustering technique, four learning patterns were identified, representing the four patterns of learning with the ebook system. The authors named the patterns according to groups as follows: "preview and diligent group," "diligent group," "efficient group," and "poor performance group." Moreover, the differences between the identified learning patterns of the students in terms of their learning outcomes and the distribution of each learning variable were investigated and discussed. Finally, the correlation between students' online learning variables and their learning outcomes are presented.

In correlating the education data with the students' self-reported data and learning patterns, Maldonado-Mahauad et al. (2018) present a bottom-up approach that mines students' behavioral patterns (process mining and clustering) using the traditional top–down approach that utilizes the validated self-reported measurement (a self-regulated learning questionnaire). In their study, three learning patterns were identified from the sequential behaviors of 3,458 online learners based on their interactions with a MOOC lecture: "sampling learners," "comprehensive learners," and "targeting learners." Furthermore, their study investigated the differences between the identified learning patterns of students in terms of their self-regulated learning (SRL) profiles and the learning outcomes of the lecture.

Compared with the literature discussed herein, the current study focused on identifying the subgroups of students with different learning patterns when learning using an ebook system in an undergraduate course. Moreover, the association between the identified subgroups of students with different patterns of learning and their learning outcomes from the course when learning with an ebook system were explored and discussed.

3. Study context and data collection

3.1. Context of the exploratory study and participants

The data used in the present study were collected from an undergraduate course called Accounting Information Systems taught at a university. A total of 113 undergraduate students enrolled in the course. These students were from the Department of Accounting. To support the instructor and students, an ebook system called BookRoll developed by Ogata et al. (2015), was used. Students enrolled in this course were allowed to study the learning material using the BookRoll system at any time. Students' online learning logs, created while interacting with BookRoll, were tracked and recorded in a database.

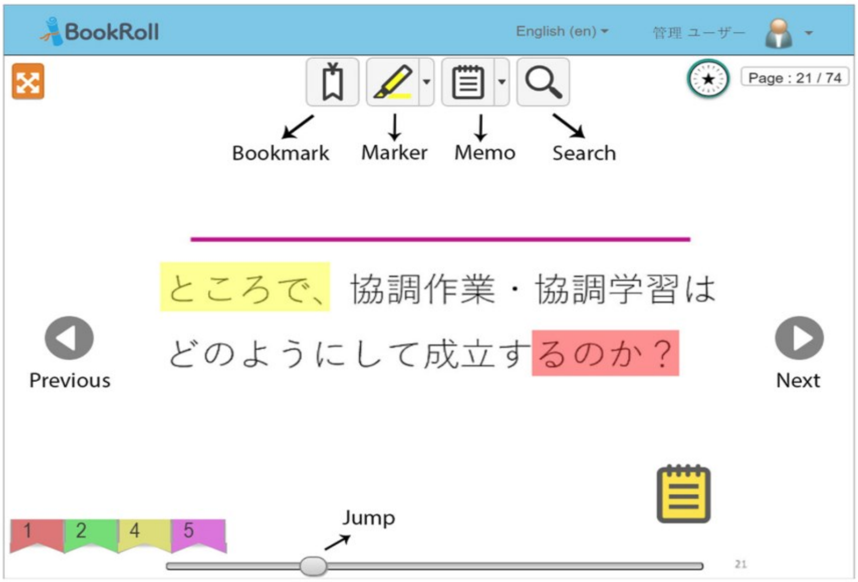


Figure 1. An example of the user interface of BookRoll.

Table 1. Example of the BookRoll learning behavioral data collected in this study

User_ID	Content_ID	Operation_Name	Operation_Date
15910	ed645f3821e	OPEN	2019/3/3 10:03:52
15910	ed645f3821e	ADD MEMO	2019/3/3 10:04:32
15923	ed645f3821e	OPEN	2019/3/3 10:07:03
15926	ed645f3821e	OPEN	2019/3/3 11:27:14
15926	ed645f3821e	NEXT	2019/3/3 11:27:20

Table 2. Number of occurrences of each BookRoll learning behavior

Learning behavioral data	Number of occurrences
OPEN	2,029
NEXT	35,988
PREV	10,167
ADD MARKER	9,125
ADD MEMO	7,284
ADD BOOKMARK	1,429
DELETE MARKER	659
DELETE MEMO	749
DELETE BOOKMARK	101
CLOSE	1,375
Total	68,906

An example of the user interface of BookRoll is displayed in Figure 1. An example of the students' learning logs collected in BookRoll is presented in Table 1. BookRoll allows users to browse the digital learning materials at any time and place after they are uploaded. Several functions are available such as page turning, marker drawing, memo taking, and page jumping. Data on the students' learning behaviors while using BookRoll are stored on its internal database. For the present study, students enrolled in the course were encouraged to use BookRoll to freely browse the digital learning material during various periods of learning, including in-class learning and out-of-class learning. The descriptions of the functions of BookRoll are detailed in Ogata et al. (2015). For this study, 68,906 ebook learning logs from 113 undergraduate students were collected using BookRoll, and the number of occurrences of each BookRoll learning behavior is given in Table 2.

3.2. Indicators

In this study, several indicators were extracted to measure the students' learning behaviors based on their uses of the ebook system BookRoll as well as their learning outcomes from the course. The collected learning behavioral data from a total of 113 undergraduate students using BookRoll is described in Table 3. Moreover, a final exam was conducted to measure the students' learning outcomes from the course. The final exam scores were determined by the course instructor at the end of the course. The collection and analysis of the indicators used to measure the students' ebook learning behaviors and their learning outcomes are detailed as follows. Indicators of learning behavior and learning outcomes from the course:

- Backtrack reading rate (BRR): Students' BRR has been proven to positively correlate with their learning outcomes from the course (Yin et al., 2019). In this study, to analyze the students' BRR, the learning behavioral data NEXT and PREV were used together to measure the BRR for the course materials throughout the course. BRR was defined as the total number of times the student reviewed the previous page divided by the total number of times the student advanced to the next page throughout the course (# of PREV /# of NEXT). For example, a BRR value of 0.5 would correspond to a case where a student reviewed the previous page 50 times and advanced to the next page 100 times in one lecture throughout the course.
- Reading time (RT): The students' RT spent has been proven in the relevant literature to positively correlate with their learning outcomes and can be used to effectively identify subgroups of students learning with an ebook system (Yin et al., 2019; Yang et al., 2019). In this study, to analyze the students' RT over an hour, time-stamp data was collected and summed up to measure students' RT with ebook learning materials throughout the course.
- Adding annotation (AN): It is suggested that the use of annotations can facilitate students' learning activities during the course by providing support through directing attention and building both internal and external connections (Du, 2004). Accordingly, Yeh and Lo (2009) demonstrate a positive correlation between the use of annotations and students' academic performance, as the students who learned with an online annotation system that allowed them to add and delete online annotations achieved better academic performance in the course than students who did not. In this study, in order to analyze the students' behaviors in terms of AN, the total volume of learning behavioral data, ADD MARKER, ADD MEMO, and ADD BOOKMARK, was summed together to measure the students' behaviors in AN related to the ebook learning materials throughout the course.
- Deleting annotation (D-AN): Yeh and Lo (2009) demonstrated a positive correlation between the tendency to delete annotations and students' academic performance in a course that was part of the same experiment as the AN study. In the present study, to analyze the students' behaviors in terms of D-AN, the total volume of learning behavioral data, DELETE MARKER, DELETE MEMO, and DELETE BOOKMARK, was summed together to measure the students' behaviors in D-AN on the ebook learning materials throughout the course.
- Learning outcome (LO): In school education, students' learning outcomes in the course or academic performances have been reported to strongly correlate with their learning engagement (Yang et al., 2020). The students' scores in the final exam issued by the course instructor were recorded at the end of the experiment. The exam comprised 40 multiple-choice items, with a perfect score of 100. Moreover, the Kuder-Richardson Formula 20 value was 0.61, showing acceptable internal consistency of the final exam (Cortina, 1993).

Table 3. Description of the collected learning behavioral data from BookRoll

Learning behavioral data	Description of the learning behavioral data
ADD MARKER	Students added a marker on the ebook learning material
ADD MEMO	Students added a memo to the ebook learning material
ADD BOOKMARK	Students added a bookmark to the ebook learning material
DELETE MARKER	Students deleted a marker on the ebook learning material
DELETE MEMO	Students deleted a memo from the ebook learning material
DELETE BOOKMARK	Students deleted a bookmark from the ebook learning material
NEXT	Students advanced to the next page of the ebook learning material
PREV	Students returned to the previous page of the ebook learning material

3.3. Clustering analysis

To analyze the ebook learning logs collected from 113 undergraduate students in the present study, we followed the analysis procedure illustrated in Figure 2. After collecting the students' learning behavioral data and extracting the indicators of learning behavior, the authors first addressed the common scale of the dataset values. All the learning behavioral data were previously normalized to a value in the range of [0, 1] by min-max normalization for further clustering analysis. Nevertheless, raw data (before data normalization) were used when presenting the analysis results detailed in the following sections.

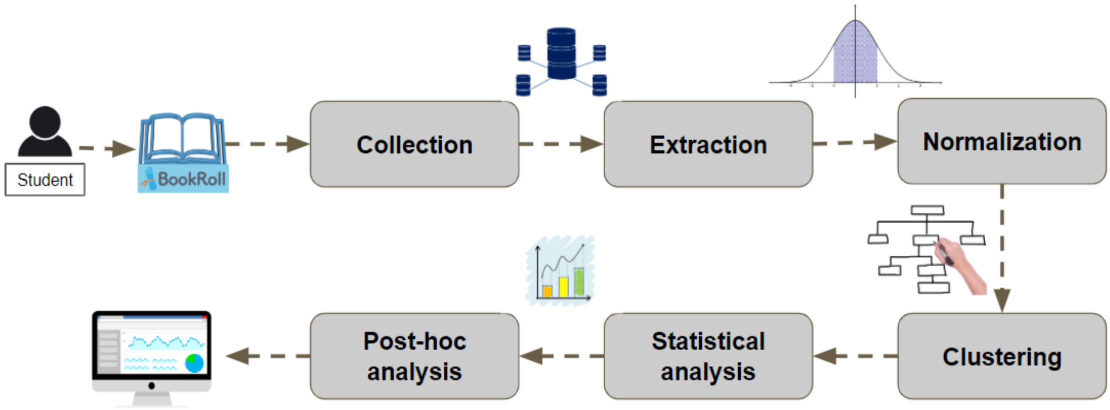


Figure 2. Analysis procedure

Next, to better understand how many subgroups of students with different learning patterns when learning with BookRoll can be identified using the four indicators of learning behavior as well as the distribution of each learning behavior between those subgroups, an agglomerative hierarchical clustering method based on Ward's method was applied. This clustering technique is recommended for identifying student subgroups in a given online learning context (Kovanović et al., 2015). The result of the cluster dendrogram displayed in Figure 3 led to the selection of three subgroups of students as the best options in this study. Thus, the students were categorized into three subgroups, which were labeled as Comprehensive learning group (CLG), Reflective learning group (RLG), and Selective learning group (SLG), based on the distribution of the four indicators of learning behavior.

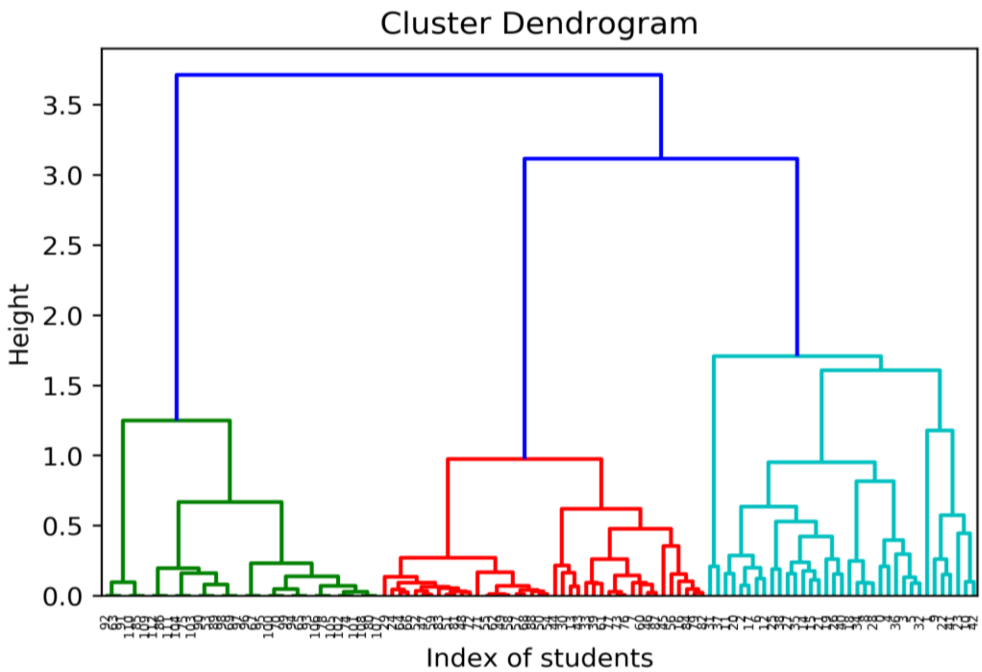


Figure 3. Dendrogram plot of clustering

4. Results

4.1. Analysis of the four indicators of learning behavior between the identified subgroups of students

After the agglomerative hierarchical clustering method was applied, the Kruskal–Wallis test was conducted to compare the identified subgroups of students in terms of the four indicators of learning behavior because the values did not have homogeneity of variance assumptions, which is required for a parametric test. Moreover, a nonparametric post-hoc test was performed using the Mann–Whitney U test to verify whether the difference between each pair of subgroups of students was statistically significant. The Kruskal-Wallis and Mann-Whitney tests have been evident with the statistical performances when the scale of data is nonparametric as well as the possibility to be applied for data analytic approaches in other face-to-face and on-line courses (Ahammed & Smith, 2019). The increasing applications of Kruskal-Wallis and Mann-Whitney U tests were identified in several research fields such as engineering applications, medicine, biology, psychology, and education (Ostertagová, Ostertag & Kováč, 2014).

According to the results listed in Table 4, significant differences were observed between the three identified subgroups of students in terms of BRR (H = 76.87, p < .001), RT (H = 66.27, p < .001), AN (H = 73.04, p < .001), and D-AN (H = 23.81, p < .001). The three identified subgroups of students are recorded as follows:

Table 4. Kruskal–Wallis test and post-hoc test (Mann–Whitney U) results of the learning behaviors between the (a) CLG, (b) RLG, and (c) SLG subgroup: median (25th percentile, 75th percentile)

	CLG (n = 35) (a)	RLG (n = 36) (b)	SLG (n = 42) (c)	H	Post-hoc test (Mann-Whitney U)
BRR	0.26 (0.17, 0.33)	0.61 (0.53, 0.72)	0.11 (0.05, 0.22)	76.87^{***}	a > c, b > a, b > c
RT	39.67 (27.61, 58.95)	14.52 (13.11, 15.29)	14.24 (12.56, 20.53)	66.27^{***}	a > b, a > c
AN	61.5 (47.45, 75.5)	16 (11.75, 24.25)	24 (13.25, 56.75)	73.04^{***}	a > b, a > c, c > b
D-AN	13 (4, 20.5)	4 (0.75, 5.75)	2 (1, 4)	23.81^{***}	a > b, a > c, b > c

Note. $^{***}p < .001$.

Comprehensive learning group (CLG): Students in this subgroup were categorized as CLG students (N = 35; 30.97% of the students) as the students in this subgroup exhibited the highest values for most of the indicators of learning behavior, representing a comprehensive way of learning as identified in Maldonado-Mahauad et al. (2018). For the BRR, CLG students exhibited higher BRR values (median = 0.26, 25th percentile = 0.17, 75th percentile = 0.33) compared with Selective learning group (SLG) students (median = 0.11, 25th percentile = 0.05, 75th percentile = 0.22). This implies that CLG students were more engaged in terms of BRR than SLG students at a statistically significant level. For RT, CLG students exhibited higher RT values (median = 39.67, 25th percentile = 27.61, 75th percentile = 58.95) compared with the Reflective learning group (RLG) students (median = 14.52, 25th percentile = 13.11, 75th percentile = 15.29) and SLG students (median = 14.24, 25th percentile = 12.56, 75th percentile = 20.53). This implies that CLG students spent more RT on learning the ebook learning materials than RLG and SLG students at a statistically significant level. For AN, CLG students exhibited higher values of AN (median = 61.5, 25th percentile = 47.45, 75th percentile = 75.5) compared with RLG students (median = 16, 25th percentile = 11.75, 75th percentile = 24.25) and SLG (median = 24, 25th percentile = 13.25, 75th percentile = 56.75). This implies that CLG students added more annotations to the ebook learning materials than RLG and SLG students at a statistically significant level. For D-AN, CLG students exhibited higher values of D-AN (median = 13, 25th percentile = 4, 75th percentile = 20.5) compared with RLG students (median = 4, 25th percentile = 0.75, 75th percentile = 5.75) and SLG students (median = 2, 25th percentile = 1, 75th percentile = 4). This implies that CLG students deleted more annotations from the ebook learning materials than RLG and SLG students at a statistically significant level.

Reflective learning group (RLG): Students in this subgroup were categorized as RLG students (N = 36; 31.86% of the students) as the students in this subgroup exhibited the highest values of BRR and higher values of D-AN compared with SLG students, representing a reflective way of learning as identified in Brinton et al. (2016). For the BRR, RLG students exhibited higher BRR values (median = 0.61, 25th percentile = 0.53, 75th percentile = 0.72) than CLG students (median = 0.26, 25th percentile = 0.17, 75th percentile = 0.33) and SLG students (median = 0.11, 25th percentile = 0.05, 75th percentile = 0.22). This implies that RLG students were more engaged in terms of BRR than CLG and SLG students at a statistically significant level. For D-AN, RLG students exhibited higher values of D-AN (median = 4, 25th percentile = 0.75, 75th percentile = 5.75) compared with SLG students (median = 2, 25th percentile = 1, 75th percentile = 4). This implies that RLG students deleted more annotations from the ebook learning materials than SLG students did at a statistically significant level.

Selective learning group (SLG): Students in this subgroup were categorized as SLG students (N = 42; 37.17% of the students) as the students in this subgroup exhibited higher values of AN compared with RLG students but exhibited the lowest values of most of the indicators of learning behavior on the other side. This represents a selective way of learning as identified in Maldonado-Mahauad et al. (2018). For AN, the SLG students had higher values of AN (median = 24, 25th percentile = 13.25, 75th percentile = 56.75) compared with RLG students (median = 16, 25th percentile = 11.75, 75th percentile = 24.25). This implies that SLG students added more annotations to the ebook learning materials than did RLG students at a statistically significant level.

4.2. Analysis of learning outcomes between the identified subgroups of students

To investigate the association between the identified subgroups of students with different patterns of learning and their learning outcomes from the course, the Kruskal–Wallis test was conducted to compare the subgroups of students in terms of their learning outcomes from the course as the values did not satisfy homogeneity of variance assumptions, which is required for a parametric test. According to the results listed in Table 5, a significant difference was observed between the three subgroups of students in terms of their learning outcomes from the course (LO; H = 14.32, p < .001). Moreover, a nonparametric post-hoc test was performed using the Mann–Whitney U test to verify whether the difference between each pair of subgroups of students was statistically significant.

The CLG students had higher values of LO (median = 83, 25th percentile = 78, 75th percentile = 85) than the SLG students did (median = 76, 25th percentile = 73, 75th percentile = 84), indicating that CLG students obtained significantly higher scores in the final exam than did SLG students. This implies that CLG students achieved better learning outcomes from the course than SLG students at a statistically significant level. Furthermore, RLG students had higher LO values (median = 84, 25th percentile = 82, 75th percentile = 86) than the SLG students (median = 76, 25th percentile = 73, 75th percentile = 84), indicating that the RLG students obtained significantly higher scores in the final exam than did SLG students. This implies that RLG students achieved better learning outcomes from the course than did SLG students at a statistically significant level.

Table 5. Kruskal–Wallis test and post-hoc test (Mann–Whitney U) results of the learning outcomes from the course between the (a) CLG, (b) RLG, and (c) SLG subgroup

Student subgroup	*N*	Median	25th percentiles	75th percentiles	*H*	Post-hoc test (Mann-Whitney U)
CLG (a)	35	83	78.13	85.85	14.32^{***}	a > c
RLG (b)	36	84.3	82.05	86.85		b > c
SLG (c)	42	76.25	73.03	84.05		

Note. $^{***}p$ < .001.

5. Discussion and practical implications

5.1. Identifying subgroups of students with different patterns of learning using an ebook system

In addressing the first research question, the current authors used an agglomerative hierarchical clustering approach based on Ward's method to identify subgroups of students with different learning patterns using an ebook system. In this study, three subgroups of students with different learning patterns were identified. To better understand the difference in the learning patterns between the subgroups of students, the authors considered the difference in each indicator of learning behavior between the subgroups using Kruskal–Wallis and Mann–Whitney U tests.

The results of these tests provide evidence that the CLG students tended to apply a comprehensive learning approach when learning with an ebook system as they showed the highest tendency among almost all the indicators of learning behavior between the three subgroups, which represents a more comprehensive and engaged learning approach that combines several learning strategies when learning with an ebook system. This finding is consistent with those of Maldonado-Mahauad et al. (2018), where the students who adopted a comprehensive learning approach usually utilized a combination of learning strategies.

RLG students tended to apply a reflective learning approach when using an ebook system as they did not spend as much RT as the CLG students did. Furthermore, RLG students did not add as many annotations to the ebook learning materials as the CLG and SLG students did. Instead, the RLG students had the highest BRR values of the three subgroups and had a higher tendency to delete annotations from ebook learning materials compared with SLG students, which represents a reflective learning strategy for an ebook system. This finding is consistent with previous studies where one subgroup of students had a higher BRR as they tended to frequently return to the previous page of the ebook learning material for review instead of going ahead to the next page (Yin et al., 2019) but with a lower tendency to add annotations to ebook learning materials (Yang et al., 2019) when learning with an ebook system. Another finding in this study demonstrates that this learning strategy also led to a higher tendency to delete annotations from ebook learning materials as the students in this subgroup typically reflected on the annotations they had made previously after returning to the previous page of the ebook learning material. This finding is consistent with the relevant literature, where the learning behavior of backtrack reading is usually connected to a reflection learning strategy of associating current knowledge with previous knowledge (Costa & Kallick, 2008).

SLG students tended to apply a selective learning approach when using an ebook system as they did not have a BRR as high as that of CLG and RLG students. Furthermore, the SLG students did not spend as much RT as the CLG students did or delete as many annotations from the ebook learning materials as the CLG and RLG students. Instead, the SLG students showed a higher tendency to add annotations to ebook learning materials compared with RLG students, which represents a selective learning strategy for using an ebook system. This finding is consistent with that of a previous study; a subgroup of students had a higher tendency to add annotations to ebook learning materials when learning with an ebook system (Yang et al., 2019).

The findings of this study in identifying three subgroups of students provide insights into the true ebook learning patterns of students. The results of the clustering analysis are consistent with the relevant literature; students can be categorized into several subgroups according to their patterns of learning with ebook systems, such as BRR, annotation usage, and RT (Yin et al., 2019; Yang et al., 2019). The students were therefore provided opportunities to accurately identify and describe how they studied and what strategies they applied to previous learning activities after receiving this information from either educators or digital learning platforms. This issue is highlighted in the literature (Zhou & Winne, 2012).

5.2. Association between the identified subgroups of students with different patterns of learning and their learning outcomes from the course when learning with an ebook system

In addressing the second research question and better understanding the association between the identified subgroups of students with different patterns of learning and their learning outcomes from the course, the current authors considered the difference in the students' learning outcomes from the course across the three subgroups of students, again using the Kruskal–Wallis and Mann–Whitney U tests.

The results of these tests demonstrate that the students in CLG achieved better learning outcomes from the course than the students in SLG at a statistically significant level, as the final exam scores of the students in CLG were significantly higher than those of the students in SLG. Therefore, it is evident that the students who adopted a comprehensive learning approach that combined several types of learning strategies achieved better learning outcomes from the course than those who adopted a selective learning approach at a statistically significant level. This finding is consistent with the relevant literature, where students who follow a comprehensive or deep learning approach achieve stronger academic performance (Bliuc et al., 2010; Ellis et al., 2008), whereas students who adopt a surface or selective learning approach achieve poorer academic performance, as such learning approach was negatively correlated with their academic performance (Richardson, Abraham & Bond, 2012).

Moreover, the results demonstrate that the RLG students achieved better learning outcomes from the course than the SLG students at a statistically significant level, as the final exam scores of the RLG students were significantly higher than those of the SLG students. Therefore, students who adopted a reflective learning approach and exhibited a higher BRR and tendency to delete annotations from ebook learning materials as well as a lower tendency to add annotations to ebook learning materials can achieve better learning outcomes from the course than students who adopted selective learning approach that involves a higher tendency to add annotations to ebook learning materials and a lower BRR and tendency to delete annotations from ebook learning materials, at a statistically significant level. This finding is consistent with the literature; students who exhibited a higher BRR achieved better learning outcomes than students with a lower BRR when using ebook systems (Yin et al., 2019). This finding can also be connected to the relevant literature; backtrack reading was positively correlated to a review learning strategy of allotting time to commit knowledge from the learning materials to students' long-term memory (Lindsey et al., 2014), thereby leading to improved student learning outcomes from the course, as proven in the current study.

The findings noted herein provide insight into the association between students' learning patterns and their learning outcomes from the course when learning with an ebook system. The researchers and educators are therefore provided opportunities to predict students' learning outcomes by analyzing their online learning behaviors and providing timely intervention for improving their learning experience, which achieves one of the goals of learning analytics as part of precision education (Lu et al., 2018). Moreover, the teachers at every education level are provided opportunities to apply the experimental results to serve as a basis for adjusting their teaching strategies or materials for achieving personalized learning in the course. Furthermore, the students are provided opportunities to adjust the learning strategies they had previously adopted to better address changing learning situations and learning goals on receiving this information from either educators or digital learning platforms; this issue has been highlighted in the previous literature (Lust, Elen & Clarebout, 2013).

5.3. Limitations

This study has several limitations that must be considered. First, the sample size of this study was 113 students. The results, although significant, may not be generalized to students from other institutions of higher education. A more general analytical model is required to suit students from different institutions using the same analytics method. Next, this study focused exclusively on identifying subgroups of students with different learning patterns using ebook learning logs. Consequently, this highlights a limitation in the types of student learning behaviors that can be identified when applying the LA and EDM approaches. Therefore, various digital learning platforms can be integrated to obtain a wider range of student learning behaviors when applying similar analytic methods in future studies. Moreover, a clustering approach was applied once to identify subgroups of students with different learning patterns. It is possible to combine the clustering approach with other techniques such as process mining or sequential data mining, which may be expected to provide deeper insights into students' learning patterns regarding an ebook system.

6. Conclusion

The rising volume of students' learning behavioral data gathered by virtual learning environments provides opportunities for mining students' patterns of learning (Yang et al., 2019). Consequently, the associations and patterns between students' learning behaviors and learning outcomes can be used to trigger a learning process and thereby reach specific goals for both learning and teaching (Reimann, 2016). However, many learning patterns and strategies cannot be easily identified from the system log data without the application of data analytics methods or data mining techniques. Hence, without a series of data analytics methods for the identification of students' learning strategies and their learning patterns using LA and EDM approaches, students often have difficulty adjusting the learning strategies they previously adopted to better address changing learning situations (Lust, Elen & Clarebout, 2013).

Thus, in this study, data analytic methods for examining the learning patterns of students while learning with an ebook system were applied. An agglomerative hierarchical clustering approach was applied to identify subgroups of students with different learning patterns when learning with an ebook system in an undergraduate course for one semester. Several subgroups were identified and categorized as Comprehensive learning group (CLG), Reflective learning group (RLG), and Selective learning group (SLG) based on the different learning strategies the students adopted, such as the higher/lower values of BRR and RT and a tendency toward AN and D-AN across ebook learning materials. Moreover, the association between the subgroups of students with different patterns of learning and their learning outcomes from the course was investigated in this study. It is therefore suggested that by applying the combined LA and EDM approaches to identify and analyze the subgroups of students learning with an ebook system, the instructor may have an opportunity to not only to improve their method of teaching during the course but also to support students in taking suitable actions with recommendations to achieve particular learning goals based on the information derived from the learning patterns of students. Moreover, the authors hope that this study will motivate other researchers to use LA and EDM approaches more often and explore their possibilities for future research. The findings of this study provide not only a detailed demonstration of applying a series of data analytics methods to identify subgroups of students with different patterns of learning throughout a semester in an undergraduate course but also offer insights into the association between students' learning patterns and outcomes from the course when learning with an ebook system, which are expected to make contributions on facilitating the practices of precision education.

Acknowledgement

This work was partially supported by JSPS Grant-in-Aid for Scientific Research (S)16H06304 and NEDO Special Innovation Program on AI and Big Data 18102059-0.

References

Ahammed, F., & Smith, E. (2019). Prediction of students' performances using course analytics data: A Case of water engineering course at the university of south Australia. *Education Sciences*, *9*(3), 245. doi:10.3390/educsci9030245

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In *Learning analytics* (pp. 61-75). New York, NY: Springer.

Bliuc, A. M., Ellis, R., Goodyear, P., & Piggott, L. (2010). Learning through face-to-face and online discussions: Associations between students' conceptions, approaches and academic performance in political science. *British Journal of Educational Technology*, *41*(3), 512-524.

Brinton, C. G., Buccapatnam, S., Chiang, M., & Poor, H. V. (2016). Mining MOOC clickstreams: Video-watching behavior vs. in-video quiz performance. *IEEE Transactions on Signal Processing*, *64*(14), 3677-3692.

Chen, N. S., Yin, C., Isaias, P., & Psotka, J. (2020). Educational big data: extracting meaning from data for smart education. *Interactive Learning Environments, 28*(2), 142-147. doi:10.1080/10494820.2019.1635395

Colvin, C., Rogers, T., Wade, A., Dawson, S., Gašević, D., Buckingham Shum, S., Karen J, N., Shirley, A., Lori, L., Gregor, K., Linda, C., & Fisher, J. (2015). *Student retention and learning analytics: A Snapshot of Australian practices and a framework for advancement.* Canberra, Australia: Office of Learning and Teaching, Australian Government.

Corrin, L., & De Barba, P. (2015). How do students interpret feedback delivered via dashboards? In *Proceedings of the fifth international conference on learning analytics and knowledge* (pp. 430-431). New York, NY: ACM.

Cortina, J. M. (1993). What is coefficient alpha? An Examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98-104. doi:10.1037/0021-9010.78.1.98

Costa, A. L., & Kallick, B. (Eds.). (2008). *Learning and leading with habits of mind: 16 essential characteristics for success*. Alexandria, VA: ASCD.

Du, M. C. (2004). *Personalized annotation management for web based learning service* (Unpublished master thesis). National Central University, Chungli, Taiwan.

Ellis, R. A., Goodyear, P., Calvo, R. A., & Prosser, M. (2008). Engineering students' conceptions of and approaches to learning through discussions in face-to-face and online contexts. *Learning and Instruction*, *18*(3), 267-282.

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5-6), 304-317.

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The Effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68-84.

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A Generic framework for learning analytics. *Educational Technology & Society*, *15*(3), 42-57.

Hu, Y. H., Lo, C. L., & Shih, S. P. (2014). Developing early warning systems to predict students' online learning performance. *Computers in Human Behavior*, *36*, 469-478.

Hwang, G. J. (2014). Definition, framework and research issues of smart learning environments-a context-aware ubiquitous learning perspective. *Smart Learning Environments, 1*(1), 4. doi:10.1186/s40561-014-0004-5

Hwang, G. J., Chu, H. C., & Yin, C. (2017). Objectives, methodologies and research issues of learning analytics. *Interactive Learning Environments, 25*(2), 143-146.

Juhaňák, L., Zounek, J., & Rohlíková, L. (2019). Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. *Computers in Human Behavior*, *92*, 496-506.

Kovanović, V., Gašević, D., Joksimović, S., Hatala, M., & Adesope, O. (2015). Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions. *The Internet and Higher Education*, *27*, 74-89.

Krumm, A. E., Waddington, R. J., Teasley, S. D., & Lonn, S. (2014). A Learning management system-based early warning system for academic advising in undergraduate engineering. *Learning analytics* (pp. 103-119). New York, NY: Springer.

Kumar, K., & Kumar, V. (2018). Advancing learning through smart learning analytics: A Review of case studies. *Asian Association of Open Universities Journal*, *13*(1), 1-12.

Lindsey, R. V., Shroyer, J. D., Pashler, H., & Mozer, M. C. (2014). Improving students' long-term knowledge retention through personalized review. *Psychological science*, *25*(3), 639-647.

Lu, O. H., Huang, A. Y., Huang, J. C., Lin, A. J., Ogata, H., & Yang, S. J. (2018). Applying learning analytics for the early prediction of students' academic performance in blended learning. *Educational Technology & Society*, *21*(2), 220-232.

Lust, G., Elen, J., & Clarebout, G. (2013). Regulation of tool-use within a blended course: Student differences and performance effects. *Computers & Education*, *60*(1), 385-395.

Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R. F., Morales, N., & Munoz-Gama, J. (2018). Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses. *Computers in Human Behavior*, *80*, 179-196.

Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). E-Book-based learning analytics in university education. In *International Conference on Computer in Education (ICCE 2015)* (pp. 401-406). Hangzhou, China: Asia-Pacific Society for Computers in Education.

Ostertagová, E., Ostertag, O., & Kováč, J. (2014). Methodology and application of the Kruskal-Wallis test. In *Applied Mechanics and Materials, 611,* 115–120. doi:10.4028/www.scientific.net/amm.611.115

Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A Systematic literature review of empirical evidence. *Educational Technology & Society*, *17*(4), 49-64.

Pardo, A., Han, F., & Ellis, R. A. (2016). Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance. *IEEE Transactions on Learning Technologies*, *10*(1), 82-92.

Reimann, P. (2016). Connecting learning analytics with learning research: The Role of design-based research. *Learning: Research and Practice*, *2*(2), 130-142.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A Systematic review and meta-analysis. *Psychological bulletin*, *138*(2), 353-387.

Romero, C., Cerezo, R., Bogarín, A., & Sánchez-Santillán, M. (2016). Educational process mining: A Tutorial and case study using moodle data sets. *Data mining and learning analytics: Applications in educational research*, 1-28.

Romero, C., López, M. I., Luna, J. M., & Ventura, S. (2013). Predicting students' final performance from participation in on-line discussion forums. *Computers & Education*, *68*, 458-472.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46*(5), 30.

Winne, P. H., & Jamieson-Noel, D. (2003). Self-regulating studying by objectives for learning: Students' reports compared to a model. *Contemporary Educational Psychology*, *28*(3), 259-276.

Wong, B. T. M. (2017). Learning analytics in higher education: An Analysis of case studies. *Asian Association of Open Universities Journal*, *12*(1), 21-40.

Yang, C. C., Chen, I. Y., Huang, A. Y., Lin, Q. R., & Ogata, H. (2020). Can self-regulated learning intervention improve student reading performance in flipped classrooms? *International Journal of Online Pedagogy and Course Design (IJOPCD)*, *10*(4), 1-13.

Yang, C. C., Flanagan, B., Akçapınar, G., & Ogata, H. (2019). Investigating subpopulation of students in digital textbook reading logs by clustering. In *Proceedings of the 9th International Conference on Learning Analytics and Knowledge (LAK'19)* (pp. 465-470). Tempe, AZ: Society for Learning Analytics Research (SoLAR).

Yang, S. J. H. (2019). Precision education: New challenges for AI in education [conference keynote]. In *Proceedings of the 27th International Conference on Computers in Education (ICCE)* (pp. XXVII-XXVIII). Kenting, Taiwan: Asia-Pacific Society for Computers in Education (APSCE).

Yeh, S. W., & Lo, J. J. (2009). Using online annotations to support error correction and corrective feedback. *Computers & Education*, *52*(4), 882-892.

Yin, C., Yamada, M., Oi, M., Shimada, A., Okubo, F., Kojima, K., & Ogata, H. (2019). Exploring the relationships between reading behavior patterns and learning outcomes based on log data from e-books: A Human factor approach. *International Journal of Human–Computer Interaction*, *35*(4-5), 313-322.

Zhou, M., & Winne, P. H. (2012). Modeling academic achievement by self-reported versus traced goal orientation. *Learning and Instruction*, *22*(6), 413-419.

From Human Grading to Machine Grading: Automatic Diagnosis of e-Book Text Marking Skills in Precision Education

Albert C. M. Yang[1*], Irene Y. L. Chen[2], Brendan Flanagan[3] and Hiroaki Ogata[3]
[1]Graduate School of Informatics, Kyoto University, Japan // [2]Department of Accounting, National Changhua University of Education, Taiwan // [3]Academic Center for Computing and Media Studies, Kyoto University, Japan // yang.ming.35e@st.kyoto-u.ac.jp // irene@cc.ncue.edu.tw // flanagan.brendanjohn.4n@kyoto-u.ac.jp // hiroaki.ogata@gmail.com
[*]Corresponding author

ABSTRACT: Precision education is a new challenge in leveraging artificial intelligence, machine learning, and learning analytics to enhance teaching quality and learning performance. To facilitate precision education, text marking skills can be used to determine students' learning process. Text marking is an essential learning skill in reading. In this study, we proposed a model that leverages the state-of-the-art text summarization technique, Bidirectional Encoder Representations from Transformers (BERT), to calculate the marking score for 130 graduate students enrolled in an accounting course. Then, we applied learning analytics to analyze the correlation between their marking scores and learning performance. We measured students' self-regulated learning (SRL) and clustered them into four groups based on their marking scores and marking frequencies to examine whether differences in reading skills and text marking influence students' learning performance and awareness of self-regulation. Consistent with past research, our results did not indicate a strong relationship between marking scores and learning performance. However, high-skill readers who use more marking strategies perform better in learning performance, task strategies, and time management than high-skill readers who use fewer marking strategies. Furthermore, high-skill readers who actively employ marking strategies also achieve superior scores of environment structure, and task strategies in SRL than low-skill readers who are inactive in marking. The findings of this research provide evidence supporting the importance of monitoring and training students' text marking skill and facilitating precision education.

Keywords: Text summarization, Marker grading, Self-regulated learning, Precision education, Text marking

1. Introduction

Precision education (Yang, 2019) is a challenge in applying artificial intelligence and machine learning techniques as well as learning analytics to enhance teaching quality and learning performance. Its goal is to identify at-risk students as early as possible and provide timely assistance through diagnosis, prediction, treatment, and prevention—specifically, it aims to identify students' learning patterns and behaviors during the learning process and predict their learning outcomes. The instructor can then provide personalized feedback or intervention to those at-risk students to prevent them from failing the course.

With advancements in information and communication technology, learning actions can be logged by a learning management system, such as Moodle (Ogata et al., 2017). Learning analytics can be used to analyze these logs—for example, the time students stay on a certain page or the markings or notes they make. Nian et al. (2019) and Yin et al. (2019) analyzed how often students used the e-reader functions such as NEXT, PREV, and MARKER to determine whether these behaviors are related to learning performance. They found that students who used the marker function tended to achieve superior learning performance. Al-khazraji (2019) observed that learners who used markers to highlight the sentences they thought were important exhibited considerably improved learning effectiveness. Yufan et al. (2020) proposed that in addition to analyzing the marking frequency, the area of marking may also be related to learning performance. However, if the content of the text being marked is not considered, marking can be overused or misused, resulting in decreased learning performance.

Measuring the content of markings reveals student's comprehension of the course. Normally, students mark the sentences or words they perceive as important during their learning process. The increasing use of e-learning systems has made it easier for instructors to observe students' learning behavior and provide relevant advice. Traditionally, the assessment process has typically been performed by instructors or teachers. However, this process is unsuitable when teaching resources are limited. To address this issue, text summarization techniques can be applied to automate the assessment process.

The adoption of marking or underlining is a metacognitive skill that enables learners to identify the essential concepts and focus on them during the review process (Van Horne et al., 2016). Similarly, self-regulated learning (SRL) is a learning process that includes multiple metacognitive skills and positively affects learning performance (Michalsky & Schechter, 2013; Siadaty et al., 2012). We speculated that learners with SRL skills are more likely to adopt and optimize the effect of the text marking strategy. For instance, students who are good at task strategies may mark the critical concepts to enhance their memory retention. Therefore, this study also evaluated the correlation between text marking and SRL.

2. Literature review

2.1. Precision education

Precision education involves four phases: diagnosis, prediction, treatment, and prevention. The process of identifying students' learning patterns and behavior, predicting learning outcomes based on the collected data, providing timely intervention, and preventing them from failing in the course is similar to the procedure when a doctor gives a patient a treatment based on their symptoms, hence the name precision education. Contrary to the traditional one-size-fits-all approach to teaching, precision education aims to provide individual students with personalized feedback and treatment according to their learning profile. Empirical studies have shown the effectiveness of precision education. For example, Lu et al. (2018) applied learning analytics early on in a blended calculus course to predict students' academic performance and identified seven critical factors affecting their performance. Hu et al. (2014) developed early-warning systems to predict at-risk students during a course in progress. They found that time-dependent variables are essential to predict student online learning performance. In this study, we explored whether the content marked by students in an e-book can predict their learning performance.

2.2. Text summarization techniques for key concept extraction

Automated text summarization has been studied since the late 1950s (Luhn, 1958). Many studies have focused on extractive summarization using statistical methods (Das & Martins, 2007). TextRank is an extractive and unsupervised text summarization technique introduced by Rada Mihalcea and Paul Tarau (Mihalcea & Tarau, 2004). It has been used to summarize meeting transcripts (Garg et al., 2009) and assess web content credibility (Balcerzak et al., 2014). Rose et al. (2010) introduced rapid automatic keyword extraction (RAKE), an unsupervised, domain-independent, and language-independent method for text summarization. They applied RAKE to a corpus of news articles to extract keywords that are essential to documents. While these studies achieved information retrieval by using traditional machine learning algorithms, researchers from Google built an unsupervised learning architecture called Bidirectional Encoder Representations from Transformers (BERT; Devlin et al., 2018). It is a deep learning model developed on top of the Transformer architecture (Vaswani et al., 2017), which beats nearly all existing models in the natural language processing (NLP) field for various tasks (Devlin et al., 2018).

2.3. SRL in learning analytics

SRL is a learning process involving cognitive and metacognitive strategies that enhance students' motivation to learn and reflect on their learning process, thereby contributing to their comprehension of studied materials (Michalsky & Schechter, 2013; Siadaty et al., 2012). Through SRL, students can come to deeply understand complex topics in the learning process (Jacobson & Archodidou, 2012; Järvelä et al., 2015; Labuhn et al., 2008). However, their behaviors and attitudes reflect SRL, which also contributes to their self-confidence (Artino & Jones, 2012; Stefanou et al., 2014).

After realizing the importance of SRL in online learning, studies have investigated whether the use of SRL strategies influences students' learning and metacognitive skills. For example, Littlejohn et al. (2016) discovered that learners' motivations and goals influence how they conceptualize the purpose of a course, which in turn influences their perception of the learning process. Similarly, Kizilcec et al. (2017) explored learners' SRL skills based on their overall course achievement, engagement in course content, and survey responses across six massive open online courses. They compared various SRL strategies and learners' actual behaviors and found that goal setting and strategy planning were associated with the attainment of personal course goals. Moreover, learners with high self-reported SRL were more likely to review the learned content.

Learning analytics have potential advantages in the examination of student SRL in online learning environments (Järvelä et al., 2016). The use of markers on e-book readers is considered a cognitive and metacognitive strategy (Van Horne et al., 2016). High-skill readers who have good metacognitive knowledge can identify and isolate essential concepts, whereas low-skill readers have difficulties in marking the most relevant information, leading to the overuse or misuse of the marking strategy. In this study, we collected data on students' self-reported SRL by asking them to complete an SRL questionnaire before the experiment. Their marker records were logged in the BookRoll database, and the learning analytics approach was applied to investigate their marking patterns and frequencies and SRL skills. We hypothesized that learners who are high-skill readers and frequent marker users have higher SRL skills. This approach may provide a new method to analyze the relationship between different groups of students stratified by reading and SRL skills.

2.4. The association between learning analytics, learning behaviors, and marking strategy

Using markers is considered a metacognitive learning behavior that positively influences reading comprehension (Van Horne et al., 2016). According to Pearson et al. (2016), the prime tasks of students when reading a textbook are to (a) focus attention and (b) engage in encoding activities in a manner that will increase the probability of understanding and retrieving the high pay-off ideas and relationships. Thus, they should identify the most relevant information in the text and test themselves to ensure that the material is understood and remembered. Nist and Hogrebe (1987) argued that students cannot remember everything they read. Thus, the use of text marking can help them identify and isolate key concepts. Specifically, college students are often required to learn and remember a large amount of information for assessments over a relatively long period. The strategy of marking or underlining the most relevant information can help them reduce the amount of information they need to review. Nist and Hogrebe (1987) further argued that instead of merely engaging in passive reading, text marking allows students to actively engage with the learning materials, which is believed to improve their memory retention. Chickering and Gamson (1987) also suggested the importance of active learning—learning that engages the student and makes them active participants in the learning process.

Studies have demonstrated the importance of marking skill in reading and learning performance. Bell and Limber (2009) noted differences in reading skills based on how much students mark, with low-skill readers marking more than high-skill readers. Nian et al. (2019) explored correlations between student reading engagement and learning outcomes and applied machine learning to predict learning outcomes. Yin et al. (2019) analyzed reading pattern behaviors, such as deleting markings or bookmarks after adding them. These findings suggest that instructors must train students on how to effectively mark a text. Many researchers have simply measured the frequency of text marking in digital textbooks without considering the content being marked, thus neglecting overmarking or mismarking by students, which can confound the results.

To overcome this limitation and investigate whether marked content has a strong correlation with learning performance, we used a method to automatically summarize the text from learning materials to calculate the marking score, which was referred to as marking quality in this study. We further measured students' reading skills using their marking score and marking frequency and evaluated whether it affects learning performance and SRL. This study addressed the following research questions:

- Can machines extract concepts that are approximate to the key concepts extracted by humans for marker grading?
- What is the relationship between marking quality, marking frequency, and learning outcomes?
- Is there a difference among students with varying levels of reading skills in learning performance and SRL?

3. Methods

3.1. Research context

A 12-week accounting course offered to undergraduate students at a university in Taiwan was the research context. The course was mandatory for students majoring in accounting but open to other majors as an elective course. The course used an e-book system, BookRoll, in which students read the e-books uploaded by the instructor. Thirteen slides were uploaded to BookRoll. BookRoll is an e-book reading system (Flanagan & Ogata, 2017) developed by Kyoto University. Students' e-book reading actions in BookRoll have been introduced in detail by Ogata et al. (2015) and Flanagan and Ogata (2018). One hundred thirty-two undergraduate students in the department of accounting took the course. Students completed a questionnaire containing SRL questions at the beginning of the course. Students took a midterm exam in the middle of the semester and a final exam at the end of the semester. All students completed the course, but two failed to complete the SRL questionnaire, and consequently, we used the data from 130 students for analysis.

3.2. Procedure

In the first week, the students completed the prequestionnaire, which comprised items related to self-regulation. The instructor then introduced the syllabus of the course and demonstrated how to use BookRoll functions, such as opening slides, using markers, and adding memos. Before each class, students were required to preview the learning materials uploaded by the instructor on BookRoll. They were also encouraged to use the various functionalities in BookRoll, such as using markers or adding memos, as their use of those interactive tools counted toward their learning activity score. Instructors can highlight the sentences they want student to pay more attention or post notes to provide furhter explanation. At weeks 8 and 12, the students took a midterm and final exam, respectively. All activities by students on the e-books were logged in the BookRoll database. We applied BERT to automatically extract the concepts from learning materials and used them as reference answers to calculate the Bilingual Evaluation Understudy (BLEU) score (Papineni et al., 2002) of student markings as their marking score. The students received a BLEU score ranging from 0 to 1 each week, and the sum of scores over 12 weeks constituted their final marking score. We also determined the frequency of using markers by collecting the total number of markings made and preserved by students at the end of the experiment. We calculated the Spearman correlation coefficient between the final marking score, marking frequency, and students' learning performance, which was measured using their midterm (30%) scores, final exam(40%) scores, and performance in learning activities during the class (30%), to investigate whether marking quality and marking frequency are correlated with learning performance.

In addition, we applied k-means clustering (Lloyd, 1982) to categorize students into four groups: high-skill readers who prefer marking, high-skill readers who do not like marking, low-skill readers who prefer marking, and low-skill readers who do not like marking. We also compared differences in students' awareness of SRL between the groups. The two features used in k-means were students' reading skills and activeness in using markers. Reading skill was measured using students' final marking score, whereas activeness in using markers was assessed using marking frequency.

3.3. Preprocessing

Automatic extraction of key concepts by using text summarization models requires text preprocessing. We applied Python's pdfminer package to convert the PDF learning materials to plain text files and extract the sentences. Both instructor and students' markers were collected using BookRoll. We removed special characters throughout and converted the text to lowercase. In addition, we ignored the deleted markings and used only the marking that were preserved by the students. Furthermore, we added two special tokens, [CLS] and [SEP], before and after the input for the BERT model, respectively. The [CLS] token in the output of BERT stores the embeddings of the text that represent its syntax and meaning. The [SEP] token serves as the separator between sentences. To simplify the preprocessing steps for BERT, the open-source transformers package developed by the Huggingface team was used.

3.4. Text summarization

We compared three text summarization techniques for automatically extracting key concepts from learning materials. For the traditional algorithms—TextRank and RAKE—the preprocessed text was passed as the input to the models and the keywords in the learning materials were extracted. TextRank is an unsupervised machine learning algorithm used to extract keywords from a text. It applies the idea of the PageRank algorithm (Page et al., 1999) developed by Google for webpage search to calculate the weights of each keyword. RAKE is an unsupervised, language-independent machine learning algorithm that extracts both keywords and key phrases. It splits the text into sentences using special characters and breaks downs each sentence at stop words. The idea behind this is that keywords usually contain multiple words and are surrounded by stop words. After extracting the candidate phrases, each candidate phrase is weighted by the co-occurrence of the words it contains. We selected the top 15 keywords based on the weights and the sentences, which include those words related to the key concepts in the materials. In this study, we adopted open-source TextRank4ZH and Rake_For_Chinese to implement these two algorithms.

BERT is a state-of-the-art technique that adopts the popular two-step transfer learning (Torrey & Shavlik, 2010) used in the NLP field. First, a general model that can understand the basic syntax of the natural language is generated during the pretraining phase, which can then be used for feature extraction or fine-tuned to perform downstream tasks. The self-attention mechanism allows BERT to understand the contextual meaning and learn the syntax structure in the text. For BERT, a pretrained model developed by the transformers team was adopted. We first passed the text of materials to the BERT encoder and extracted the last hidden state of the [CLS] token to acquire a 768-dimension embedding representing the key concepts of materials. To overcome the limit of the number of words that can be sent to the model at one time, we divided the text by pages, retrieved the embeddings of each page, and summed them to acquire the embeddings of the complete text. Next, we applied the same procedure to obtain the embeddings of each sentence in the text. Finally, we calculated the cosine distance between the embeddings of materials and the embeddings of each sentence in the vector space to measure whether the sentence is similar to the key concepts in the materials; we selected the sentences closest to the materials as key concepts.

To compare the performance of the three models, we used the markers provided by the instructor as the reference answer to evaluate the quality of summaries by the machine using BLEU 1, BLEU 2, BLEU 3, BLEU 4 (Papineni et al., 2002), and METEOR (Denkowski & Lavie, 2014) scores. BLEU is a precision-based measure for evaluating a hypothesis translation of a text to reference translations. BLEU-n is a variant of the BLEU score that applies up to a specified n-gram for counting co-occurrences. METEOR is a recall-oriented metric that evaluates the generated sentences using stemming and synonymy matching along with standard exact word matching. Both BLEU and METEOR scores range from 0 to 1. The higher the score is, the better the machine-generated sentences are. In this study, we calculated four scores for the summaries generated by different models every week and summed the scores as the performance of each model.

4. Results

4.1. Analysis of machine grading and human grading

Table 1 presents the scores of each model in different metrics. BERT performed the best in all metrics except for BLEU-4. Therefore, we adopted BERT to extract key concepts from the learning materials, with these extractions serving as the reference answer for marker grading in the following experiment. Considering that most keywords in the key concepts and student markings contained only one or two words, BLEU-1 was chosen as the metric when grading markers.

Table 1. Evaluation of TextRank, RAKE, and BERT

	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR
TextRank	7.18	5.39	4.21	**3.16**	.25
RAKE	7.4	5.48	4.13	3.12	.26
BERT	**7.98**	**5.97**	**4.31**	3.03	**.43**

Table 2. Cohen's kappa coefficients of different grading results

Human	Machine		
	< 0.2	0.2–0.8	> 0.8
< 0.2	17	0	0
0.2–0.8	5	43	0
➢ 0.8	0	16	7

Table 3. Classification reports of the BERT model

Marking score	Precision	Recall	*F1*
< 0.2	0.77	1.0	0.87
0.2–0.8	0.73	0.9	0.80
> 0.8	1.00	0.3	0.47

Student markings were graded, and the sum of marking scores was used as their final marking score. Before analyzing the correlation between marking score and learning performance, the machine-grading results were compared with human-graded marking to ensure its reliability by calculating Cohen's kappa coefficient, which is often used to measure interrater agreement. Table 2 presents the grading results for machines and humans. The Cohen's kappa coefficient was 0.57, indicating a moderate agreement between the two grading results. The Spearman correlation between the two grading results (rs = 0.84^{***}, $p < .001$) indicated that two grading approaches were highly correlated. Table 3 presents the classification reports of the BERT model. The model achieved the best *F1* score of 0.87 when the marking score was < 0.2 and the worst *F1* score of 0.47 when the marking score was > 0.8, meaning that our model is good at identifying bad markers but may sometimes generate low scores for students who actually mark well. To make sure the marking score used in this study would reflect students' actual performance, the machine grading results were cross-examined by the instructor. If the instructor did not agree with the scores, the scores were adjusted. On the basis of our results, we conclude that machine grading can be used to automatically grade student markings but still requires assistance from humans for a specific group of students.

4.2. Analysis of marking frequency and quality

To compare differences between high and low achievement groups in the score and number of markings, the Kruskal–Wallis H-test was performed. The Shapiro–Wilk test of marking score and marking frequency indicated a nonnormal distribution of these variables. We divided the students into high (n = 44) and low achievement (n = 44) groups by taking the upper and lower thirds of the learning performance from the 130 participants. The high achievers did not perform differently in marking score ($H = 0.07$, $p > .05$) and marking frequency ($H = 0.93$, $p > .05$) than low achievers (Table 4). Because no clear linear relationship was observed between marking score, marking frequency, and learning performance, the Spearman correlation coefficient was calculated to measure the relationship among them (Table 5). No correlation was noted between marking score and learning performance (rs = −0.01, p >.05). A weak but nonsignificant correlation was noted between marking frequency and learning performance (rs = 0.12, p >.05). Finally, the marking score was highly correlated with marking frequency (rs = 0.68^{***}, $p < .001$). However, this result does not indicate that students can get a high marking score by simply marking more. In Figure 1, when the number of markings is below a threshold of 200, the marking score improves rapidly as the students make more markings. However, once the number exceeds 200, the improvement decreases drastically. We further averaged the marking scores of all students through the semester to investigate whether their ability to identify the key concepts from the learning materials varied in different periods (Figure 2). From Weeks 1 to 6, the marking score was relatively moderate and stable, meaning that students may not have found the motivation to use the marking strategy or were still unfamiliar with the system. The average of marking scores reached the peaks in Weeks 7 and 11, indicating that the students were more familiar with the system and were marking more frequently before the midterm and final exams in Weeks 8 and 12. In Week 9, no new slides were uploaded because the instructor was reviewing the midterm exam with students. To answer the research question, we conclude that, consistent with previous studies (Fowler & Barker, 1974; Hoon, 1974; Idstein & Jenkins, 1972), neither marking frequency nor marking score has a significant correlation with learning performance. However, a strong relationship exists between marking score and marking frequency, which leads into the discussion of our third research question.

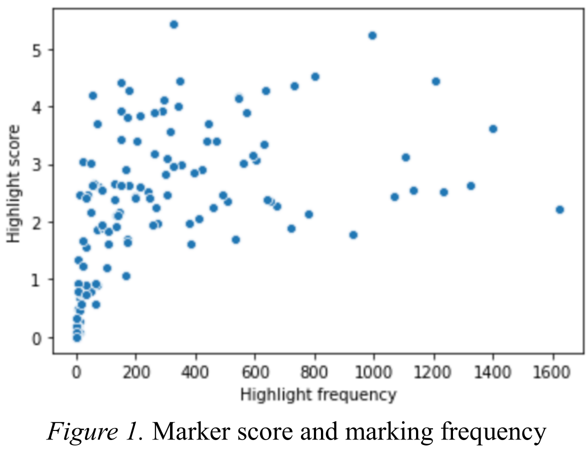


Figure 1. Marker score and marking frequency

Table 4. Results of marking score and marking frequency

	High achievement			Low achievement			
	N	*M*	*SD*	*N*	*M*	*SD*	*H*
Marking score	44	2.40	1.43	44	2.31	1.31	0.07
Marking frequency	44	271.0	439.76	44	152.5	301.25	0.93

Table 5. Correlations between marking score, marking frequency, and learning performance

	1.Marking score	2.Marking frequency	3.Learning performance
1.Marking score	-	0.68^{***}	−0.01
2.Marking frequency		-	0.12
3.Learning performance			-

Note. $^{***}p < .001$.

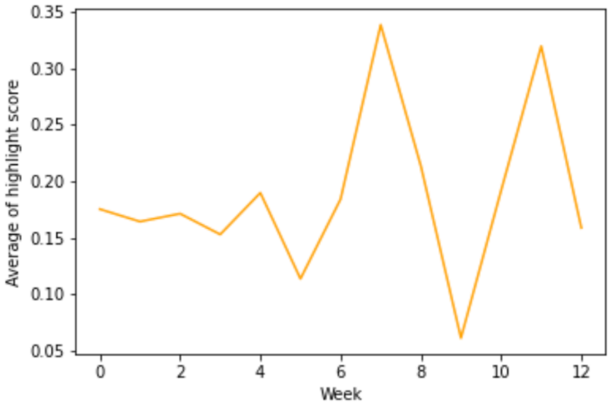


Figure 2. Average marking score throughout the semester

4.3. Analysis of learning performance

To determine whether different reading skills influence students' learning performance and SRL, we used k-means clustering to classify students into four groups: inactive marker users who are high- (Group A, $N = 60$) or low-skill readers (Group B, $N = 41$) and active marker users who are high- (Group C, $N = 27$) or low-skill readers (Group D, $N = 2$). Because Group D only contained two samples, it was not considered in the following test. The Shapiro–Wilk test results revealed a value of 0.97 ($p < .05$), indicating that this sample did not show a normal distribution.

The Kruskal–Wallis H-test was employed to evaluate students' learning achievement in the different groups. Table 6 shows the results of the learning achievement according to final grades. The medians and standard deviations were 81.0 and 7.79, respectively, for Group A; 83.0 and 7.63, respectively, for Group B; and 85.0 and 10.14, respectively, for Group C. The final grades of the three groups were not significantly different ($H = 4.13$, $p > .05$). However, the post hoc Conover test indicated that the learning performance of Group C was significantly higher than that of Group A. This implies that high-skill readers who prefer marking are more likely to also achieve better final grades than other high-skill readers who do not prefer marking.

Table 6. The Kruskal test result of learning performance of three groups

Group	*N*	*M*	*SD*	*H*	Post hoc test (Conover)
A	60	81.0	7.79	4.13	
B	41	83.0	7.63		
C	27	85.0	10.14		$C > A^{*}$

Note. $^{*}p < .05$.

4.4. Analysis of self-regulation

The SRL questionnaire comprised items regarding six aspects: goal setting, environment structure, task strategies, time management, help-seeking, and self-evaluation. We summed the scores of the six aspects to represent the total SRL score. Table 7 shows the results of the students' self-regulation along with the scores of the six aspects of SRL across the three groups. The Kruskal–Wallis H-test indicated a nonsignificant effect on students' self-regulation in all groups ($H = 4.87$, $p > .05$). The medians and standard deviations of Groups A, B, and C were 122.0 and 20.11, 118.0 and 17.42, and 131.0 and 18.95, respectively. However, the post hoc test indicated a significant difference between Groups B and C in self-regulation, implying that high-skill readers who prefer marking exhibit a better awareness of self-regulation than low-skill readers who do not prefer marking.

To further investigate the students' awareness of their self-regulation in each aspect, the Kruskal test was employed again. As presented in Table 7, no significant difference was observed in goal setting (H = 1.78, p > .05), help-seeking (H = 1.54, p > .05), or self-evaluation (H = 2.06, p > .05). However, a significant difference was noted between the three groups in environment structure (F = 6.59, p < .05). The Conover post hoc test indicated that Group C obtained a significantly higher score in the awareness of environment structure (N = 27, M = 24.0, SD = 2.81) than Group B (N = 41, M = 22.0, SD = 3.30). A significant difference in task strategies (H = 7.32, p < .05) was detected between the three groups. Group C performed significantly better than Groups A and B. Also, the time management scores between the three groups were significantly different (F = 6.13, p < .05). Group C obtained a significantly higher score than Group A.

Table 7. Kruskal–Wallis test results of SRL in the three groups

Dimension	Group	*N*	*M*	*SD*	*H*	Post hoc test (Conover)
SRL	A	60	122.0	20.11	4.87	
	B	41	118.0	17.42		
	C	27	131.0	18.95		C > B*
Goal setting	A	60	26.5	4.67	1.78	
	B	41	27.0	3.96		
	C	27	28.0	3.68		
Environment structure	A	60	24.9	3.38	6.59*	
	B	41	22.0	3.30		
	C	27	24.0	2.81		C > B*
Task strategies	A	60	19.0	3.68	7.32*	
	B	41	18.0	3.68		
	C	27	21.0	3.61		C > A*; C >B**
Time management	A	60	13.0	3.14	6.13*	C > A*
	B	41	14.0	3.20		
	C	27	15.0	3.11		
Help-seeking	A	60	20.0	4.17	1.54	
	B	41	20.0	3.75		
	C	27	21.0	4.12		
Self-evaluation	A	60	20.0	3.66	2.06	
	B	41	20.0	3.33		
	C	27	21.0	3.87		

Note. $^{**}p$ < .01, $^{*}p$ < .05.

5. Discussion

5.1. Automatic concept extraction and marker analysis of learning performance

Research question 1: Can machines extract concepts that are approximate to the key concepts extracted by humans for marker grading?

Consistent with the previous study (Devlin et al., 2018), BERT outperformed other text summarization models in the natural language understanding task in the present study. The text input to BERT included slides made by the instructor. Most of the learning materials consisted of incomplete sentences or phrases in bullet points rather than a passage or paragraphs that would include a contextual meaning. Bidirectional encoding of each sentence allowed BERT to output the embeddings containing information on syntax and semantic meanings and choose the most relevant sentences in the vector space.

Using BERT-generated summarization to evaluate whether students mark key concepts consistent with the instructor's answers, we found our model generated decent results: most students who are poor or moderate at marking can be graded correctly, but students who are good at marking may sometimes receive only moderate scores (0.2–0.8). Overall, the machine can help identify students who are unable to mark the key concepts, and then, the instructor can focus on helping them better identify the relevant concepts. This finding facilitates execution of the diagnosis phase in precision education.

Research question 2: What is the relationship between marking quality, marking frequency, and learning performance?

The finding that the content being marked and the use of the marker function did not vary greatly between high and low achievers indicates that marking score or frequency alone may not determine success in online learning courses, consistent with previous findings (Fowler & Barker, 1974; Hoon, 1974; Idstein & Jenkins, 1972; Oi et al., 2017). Nevertheless, the text marking strategy may still influence learning success. The instructor needs to guide students in how to use the appropriate text marking strategy in online learning, as many students may not be familiar with the use of many functions in the e-book system. The effectiveness of marking or underlining can be optimized when students are trained on how to use them (Nist & Simpson, 1988; Yue et al., 2015). Therefore, we recommend that instructors introduce the marker function and the text marking strategy at the beginning of the course or encourage students to use this strategy and provide students with the concepts marked by them for comparison. By achieving this, text marking skills can still become a predictor of learning performance insofar as it reflects students' metacognitive skills, which correlate with learning performance (Van Horne et al., 2016).

5.2. Analysis of reading skill, activeness of marking, learning performance, and SRL

Research question 3: Is there a difference among groups with varying levels of reading skills in learning performance and SRL?

We classified students into four groups by using k-means clustering and found that high-skill readers who prefer marking exhibited significantly higher learning performance than high-skill readers who do not prefer marking. Marking is most effective when the reader has maximum faith that the marker can discriminate between essential material and trivia (Fowler & Barker, 1974). Active marker users recognize that marking can help them identify and isolate the most relevant information for later review, thereby enhancing their understanding and long-term memory (Annis & Davis, 1978; Chickering & Gamson, 1987; Nist & Hogrebe, 1987; Rickards & August, 1975).

Furthermore, high-skill readers who prefer marking exhibited a significantly higher self-reported SRL than did low-skill readers who did not prefer marking. This suggests that students with better SRL, metacognitive strategies, and motivation believe that marking is beneficial during their learning process, which leads to better use of this strategy. We further explored students' awareness of SRL in each aspect and found that high-skill readers who are active in marking performed significantly better in environment structure than did low-skill readers who do not prefer marking. Students who can find the environment for them to focus on studying or who are good at using different learning tools during their learning process can benefit from the marker function in e-books. We also found that students who are both high-skill readers and active marker users exhibited significantly better task strategies than other groups. Students who exhibit highly developed task strategies use the marker function to identify the relevant information and ignore the trivia, which enhances their reading efficiency (Nist & Simpson, 1988). Finally, high-skill readers who are active in marking exhibited significantly better time management than did high-skill readers who do not prefer marking. Students who manage their time well prefer to apply the text marking strategy during their learning, as they believe this strategy can improve their reading and learning efficiency (Nist & Simpson, 1988). This finding suggests that the combination of marking quality and marking frequency can be an indicator of students' learning performance and SRL. The instructor can provide timely interventions in terms of training on text marking skills.

6. Conclusion

Our findings provide several contributions to analysis of text marking analysis on an e-book system. First, most of the learning material in this study was PDF slides made by instructors. They extracted the passages or sentences they thought were relevant to the core of the course. These texts were often sentences without contextual meaning or phrases, which is different from a traditional textbook. To our understanding, most research on text summarization has used textbooks, papers, or datasets for competition, whereas our study extracted concepts from slides made by instructors.

Second, this study applied an advanced tool, BERT, to extract concepts from learning materials and use them as reference answers to automatically grade student markings, whereas studies have mainly explored the number, area, or pattern of markings (Yufan et al., 2020). According to Memory (1983) and Meyer et al. (1980), only high-skill readers effectively identify the most relevant materials. High-skill readers pay attention to the essential concepts in the text to a greater degree than low-skill readers do (Lorch & Pugzles-Lorch, 1985). Therefore, we measured not only marking frequency but also students' marking scores to assess their reading skills.

Third, this study revealed the correlation between reading skills, learning performance, and SRL. The combination of marking quality and marking preference has not been used previously. According to our findings, mere frequency of marking did not guarantee high marking scores. Therefore, we considered both indicators when measuring student learning performance and SRL and found that high-skill readers who are active in marking perform significantly better in learning performance, SRL, environment structure, task strategies, and time management.

Finally, the findings of this research provide insights for instructors, students, and researchers in the field of precision education. Instructors can use the proposed system to track students' text marking patterns, predict at-risk students through their marking patterns, and provide timely personalized feedback for individual students. Students can be kept aware of their text marking ability through the system. Weekly monitoring of their marking score and marking frequency enables them to determine whether their learning strategy is effective and whether they should improve their reading skills. Finally, as the current study emphasizes the diagnosis and prediction phase in precision education, future studies can focus on treatment and prevention using text marking patterns—for example, whether students improve their text marking patterns after seeing an e-book in which the key concepts have been marked by the instructor. To summarize, our findings can help instructors and students track text marking skills and demonstrate that the conceptualization of automatically grading markers is effective in the precision education field.

This study has a few limitations. First, this experiment was conducted in an accounting course that involves many terms and theories that require memory retention. Thus, marking may benefit students when a large amount of information needs to be remembered. Using the marker function allows students to filter out irrelevant information. However, whether marking is as effective when students engage in a course where computing and logic are used more frequently remains unclear. Second, the students were not given any instructions regarding marking strategies in this study during their learning. It has been shown that students well trained in marking performed better in learning performance and learning efficiency than those not trained in marking. Thus, students' marking scores can positively influence learning performance after they receive training; thus, as students improve their ability to identify key concepts, their learning performance can be enhanced.

In summary, our study leveraged the advanced text summarization model to automatically grade student markings and analyzed the effectiveness of text marking strategy on learning performance and SRL. Leutner et al. (2007) stated that it is worth training students in specific metacognitive learning strategies and teaching them to examine and regulate their use of those strategies in learning. Future studies should explore whether teaching students to mark on e-books and SRL strategies affects their learning performance.

Acknowledgement

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (B)20H01722, JSPS Grant-in-Aid for Scientific Research (S)16H06304 and NEDO Special Innovation Program on AI and Big Data 18102059-0.

References

Al-khazraji, A. (2019). Analysis of discourse markers in essays writing in ESL classroom. *International Journal of Instruction, 12*(2), 559-572.

Annis, L., & Davis, J. K. (1978). Study techniques: Comparing their effectiveness. *The American Biology Teacher, 40*(2), 108-110.

Artino Jr, A. R., & Jones II, K. D. (2012). Exploring the complex relations between achievement emotions and self-regulated learning behaviors in online learning. *The Internet and Higher Education, 15*(3), 170-175.

Balcerzak, B., Jaworski, W., & Wierzbicki, A. (2014). Application of textrank algorithm for credibility assessment. In *Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)* (Vol. 1. pp. 451–454). doi:10.1109/WI-IAT.2014.70

Bell, K. E., & Limber, J. E. (2009). Reading skill, textbook marking, and course performance. *Literacy research and instruction, 49*(1), 56-67.

Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *AAHE bulletin, 3*, 7.

Das, D., & Martins, A. (2007). A Survey on automatic text summarization. literature survey for language and statistics. *II Course at CMU*.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805.*

Denkowski, M., & Lavie, A. (2014). Meteor universal: Language specific translation evaluation for any target language. In *Proceedings of the ninth workshop on statistical machine translation* (pp. 376-380). Baltimore, MD: Association for Computational Linguistics.

Flanagan, B., & Ogata, H. (2017). Integration of learning analytics research and production systems while protecting privacy. In *Proceedings of the 25th International Conference on Computers in Education* (pp. 333-338). Christchurch, New Zealand: Asia-Pacific Society for Computers in Education (APSCE).

Flanagan, B., & Ogata, H. (2018). Learning analytics infrastructure for seamless learning. In *Proceedings of the 8th International Conference on Learning Analytics & Knowledge (LAK18).*Retrieved from http://hdl.handle.net/2433/233071

Fowler, R. L., & Barker, A. S. (1974). Effectiveness of highlighting for retention of text material. *Journal of Applied Psychology, 59*(3), 358-364. doi:10.1037/h0036750

Garg, N., Favre, B., Reidhammer, K., & Hakkani-Tür, D. (2009). Clusterrank: A Graph based method for meeting summarization. In *Tenth Annual Conference of the International Speech Communication Association* (pp. 1499-1502). Brighton, United Kingdom: International Speech Communication Association.

Hoon, P. W. (1974). Efficacy of three common study methods. *Psychological Reports, 35*(3), 1057-1058.

Hu, Y. H., Lo, C. L., & Shih, S. P. (2014). Developing early warning systems to predict students' online learning performance. *Computers in Human Behavior, 36*, 469-478.

Idstein, P., & Jenkins, J. R. (1972). Underlining versus repetitive reading. *The Journal of Educational Research, 65*(7), 321-323.

Jacobson, M. J., & Archodidou, A. (2012). The Knowledge mediator framework: Toward the design of hypermedia tools for learning. In *Innovations in Science and Mathematics Education* (pp. 128-172). New York, NY: Routledge.

Järvelä, S., Kirschner, P. A., Panadero, E., Malmberg, J., Phielix, C., Jaspers, J., Koivuniemi, M., & Järvenoja, H. (2015). Enhancing socially shared regulation in collaborative learning groups: designing for CSCL regulation tools. *Educational Technology Research and Development, 63*(1), 125-142.

Järvelä, S., Malmberg, J., & Koivuniemi, M. (2016). Recognizing socially shared regulation by using the temporal sequences of online chat and logs in CSCL. *Learning and Instruction, 42*, 1-11.

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & Education, 104*, 18-33.

Labuhn, A. S., Boegeholz, S., & Hasselhorn, M. (2008). Fostering learning through stimulation of self-regulation in science lessons. *Zeitschrift fur Padagogische Psychologie, 22*(1), 13-24.

Leutner, D., Leopold, C., & den Elzen-Rump, V. (2007). Self-regulated learning with a text-highlighting strategy. *Zeitschrift für Psychologie/Journal of Psychology, 215*(3), 174-182.

Littlejohn, A., Hood, N., Milligan, C., & Mustain, P. (2016). Learning in MOOCs: Motivations and self-regulated learning in MOOCs. *The Internet and Higher Education, 29*, 40-48.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory, 28*(2), 129-137.

Lorch, R. F., & Lorch, E. P. (1985). Topic structure representation and text recall. *Journal of Educational Psychology, 77*(2), 137.

Lu, O. H. T., Huang, A. Y. Q., Lin, A. J. Q., Ogata, H., & Yang, S. J. H. (2018). Applying learning analytics for the early prediction of students' academic performance in blended learning. *Educational Technology & Society, 21*(2), 220–232.

Luhn, H. P. (1958). The Automatic creation of literature abstracts. *IBM Journal of research and development*, *2*(2), 159-165. doi:10.1147/rd.22.0159

Memory, D. M. (1983). Main idea prequestions as adjunct aids with good and low-average middle grade readers. *Journal of Reading Behavior, 15*(2), 37-48.

Meyer, B. J., Brandt, D. M., & Bluth, G. J. (1980). Use of top-level structure in text: Key for reading comprehension of ninth-grade students. *Reading research quarterly, 16*(1), 72-103. doi:10.2307/747349

Mihalcea, R., & Tarau, P. (2004, July). Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing* (pp. 404-411). Barcelona, Spain: Association for Computational Linguistics.

Michalsky, T., & Schechter, C. (2013). Preservice teachers' capacity to teach self-regulated learning: Integrating learning from problems and learning from successes. *Teaching and Teacher Education, 30*, 60-73.

Nian, M. W., Lee, Y. H., & Wu, J. Y. (2019). Using machine learning to explore the associations among e-reader operations and their predictive validity of learning performance. In *International Conference on Learning Analytics & Knowledge (LAK19)* (pp. 1-6). Tempe, AZ: Association for Computing Machinery.

Nist, S. L., & Hogrebe, M. C. (1987). The Role of underlining and annotating in remembering textual information. *Literacy Research and Instruction, 27*(1), 12-25.

Nist, S. L., & Simpson, M. L. (1988). The Effectiveness and efficiency of training college students to annotate and underline text. *National Reading Conference Yearbook, 37,* 251–257.

Ogata, H., Oi, M., Mohri, K., Okubo, F., Shimada, A., Yamada, M., & Hirokawa, S. (2017). Learning analytics for e-book-based educational big data in higher education. In *Smart Sensors at the IoT Frontier* (pp. 327- 350). doi:10.1007/978-3-319-55345-0_13

Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). E-Book-based learning analytics in university education. In *International Conference on Computer in Education (ICCE 2015)* (pp. 401-406). Hangzhou, China: Asia-Pacific Society for Computers in Education.

Oi, M., Okubo, F., Taniguchi, Y., Yamada, M., & Konomi, S. (2017). Effects of prior knowledge of high achievers on use of e-book highlights and annotations. In *Proceedings of the 25th International Conference on Computers in Education, ICCE 2017 - Main Conference Proceedings* (pp. 682-687). Christchurch, New Zealand: Asia-Pacific Society for Computers in Education.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the web*. Technical Report. Stanford InfoLab.

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics* (pp. 311-318). Philadelphia, PA: Association for Computational Linguistics.

Pearson, P. D., Kamil, M. L., Mosenthal, P. B., & Barr, R. (Eds.). (2016). *Handbook of reading research*. New York, NY: Routledge.

Rickards, J. P., & August, G. J. (1975). Generative underlining strategies in prose recall. *Journal of Educational Psychology, 67*(6), 860-865. doi:10.1037/0022-0663.67.6.860

Rose, S., Engel, D., Cramer, N., & Cowley, W. (2010). Automatic keyword extraction from individual documents. *Text Mining: Aapplications and Theory, 1,* 1-20.

Siadaty, M., Gasevic, D., Jovanovic, J., Pata, K., Milikic, N., Holocher-Ertl, T., Jeremić, Z., Ali, L., Giljanović, A., & Hatala, M. (2012). Self-regulated workplace learning: A Pedagogical framework and semantic web-based environment. *Educational Technology & Society, 15*(4), 75-88.

Stefanou, C., Lord, S. M., Prince, M. J., & Chen, J. C. (2014). Effect of classroom gender composition on students' development of self-regulated learning competencies. *International Journal of Engineering Education, 30*(2), 333-342.

Torrey, L., & Shavlik, J. (2010). Transfer learning. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques* (pp. 242-264). Hershey, PA: IGI global. doi:10.4018/978-1-60566-766-9.ch011.

Van Horne, S., Russell, J. E., & Schuh, K. L. (2016). The Adoption of mark-up tools in an interactive e- textbook reader. *Educational Technology Research and Development, 64*(3), 407-433.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*, 5998-6008.

Yang, S. J. H. (2019). Precision education: New challenges for AI in education [conference keynote]. In *Proceedings of the 27th International Conference on Computers in Education (ICCE)* (pp. XXVII-XXVIII). Kenting, Taiwan: Asia-Pacific Society for Computers in Education (APSCE).

Yin, C., Yamada, M., Oi, M., Shimada, A., Okubo, F., Kojima, K., & Ogata, H. (2019). Exploring the relationships between reading behavior patterns and learning outcomes based on log data from e-books: A Human factor approach. *International Journal of Human-Computer Interaction*, *35*(4-5), 313-322.

Yue, C. L., Storm, B. C., Kornell, N., & Bjork, E. L. (2015). Highlighting and its relation to distributed study and students' metacognitive beliefs. *Educational Psychology Review, 27*(1), 69-78.

Yufan, X., Xuewang, G., Li, C., Satomi, H., Yuta, T., Hiroaki, O., & Yamada, M. (2020). Can the area marked in ebook readers specify learning performance. In *Proceedings of the 10th International Conference on Learning Analytics & Knowledge (LAK20)* (pp. 638-648). Frankfurt, Germany: ResearchGate.

Feng, X., & Yamada, M. (2021). An Analytical Approach for Detecting and Explaining the Learning Path Patterns of an Informal Learning Game. *Educational Technology & Society*, *24* (1), 176-190.

An Analytical Approach for Detecting and Explaining the Learning Path Patterns of an Informal Learning Game

Xuanqi Feng[1*] and Masanori Yamada[2]
[1]Graduate School of Human-Environment Studies, Kyushu University, Japan // [2]Faculty of Arts and Science, Kyushu University, Japan // dukefeng@mark-lab.net // mark@mark-lab.net
[*]Corresponding author

ABSTRACT: It is challenging to utilize learning analytic technologies to examine gameplay log data for game-embedded assessment in the field of game-based learning. Analytical approaches based on a new perspective focusing on complicated contextual data are imperative in the current scenario. A relatively new concept called precision education, which focuses on individual learning and provides personalized and timely intervention to different learners, can be regarded as a new perspective for game-based learning. Additionally, the order of knowledge acquisition in the learning environment is a kind of learning path extracted from the contextual information of in-game behavior logs. Therefore, in this study, the authors propose a new analytical approach to identify learning path patterns and elucidate the features of these patterns for an educational game they developed. The statistical analysis shows that learners with diverse learning path patterns have different learning effects, suggesting that the learning path is an important factor in precision education. The practice of using the explanation method to examine the proposed approach can help us understand learners' knowledge acquisition and provide evidence for enhancing the accuracy of precision education and improving the quality of the educational game. The findings are expected to contribute to both game-based learning and precision education.

Keywords: Learning analytics, Game-based learning, Learning path, Informal learning, Precision education

1. Introduction

Game-embedded assessment has become the mainstream method for evaluating game-based learning. This approach assesses learners' performance by tracing and analyzing behavioral process data (Shute, 2011). It is challenging to develop game-embedded assessment utilizing methods of data science such as learning analytics. Although this topic has been studied for nearly 10 years, analytics based on new perspectives is imperative to make the assessment more effective (Kim & Ifenthaler, 2019). Precision education, which is a relatively new concept that focuses on providing an individualized learning experience and timely intervention in learning environments (Hart, 2016), can be considered a new perspective for game-based assessment. Meanwhile, the learning path with regard to learning content might be an influential factor in precision education and can be considered a new analyzed data type in game-embedded assessment. Accordingly, in this study, we propose an analytical approach with regard to the learning path patterns of an informal learning game from the perspective of precision education. The findings are expected to contribute to both game-based learning and precision education.

1.1. Game-embedded assessment for informal learning games

Although digital games are considered a form of entertainment, several studies have shown that these games have the capability to increase intrinsic learning motivation or support cognitive process (Klein & Freitag, 1991). Further, cognitive and learning scientists claim that certain features of the games, such as high interactivity and the provision of immediate feedback, can also aid effective learning environments (Shute & Ke, 2012). Therefore, digital games are considered to have the potential to become attractive and effective learning environments. This approach of using digital games to build learning environments is called "digital game-based learning" (DGBL). Additionally, the assessment of learning effects in game-based learning environments is an important factor in the field of DGBL. However, this factor cannot be comprehended easily, because learners' interactions in games are often variable and complicated, making the evaluation of performance difficult. Since playing games has a strong trait of spontaneity, many educational games have been developed for informal learning (Ke, 2016), which suggests that learning is the result of outside-of-school activities with spontaneity and less rigid curricula (Werquin, 2007). In these cases, there are no observers of learning activities. External assessments such as pre- and post- tests are feasible, but information about behavioral data at the process level is lacking. In such a situation, the learning activity becomes a "black box," and the ability to understand how players learn from the game, explain experimental results, and utilize the results of assessments to improve learning environments is limited (Loh, 2011).

To address the problem of assessment in such contexts, an approach called "game-embedded assessment" has been proposed that seeks to assess a learner's performance by tracing and analyzing behavioral process data. Since process log data can be collected stealthily even if learning activities occur in an environment without any educator observer, concerns that "learning is a black box" can be avoided (Shute, 2011). Owing to the advantages of game-embedded assessment, it has been considered the mainstream method. However, given the complexity of in-game interaction, it is challenging to analyze the collected log data to realize an effective game-embedded assessment. For the analysis, data science and technology are usually needed, and learning analytics (LA) is considered relevant since it involves "using various data such as logs about learning to clarify education and learning environment improvement" (Ifenthaler, 2015). The in-game log data analysis method has been studied for nearly 10 years, with the research framework developed from traditional evidence-centered design to the application of LA technologies (Kim & Ifenthaler, 2019). As for recent researches, Akram et al. (2018) proposed an analytical framework based on the long short-term memory network to build an automatic game-embedded assessment. Furthermore, Henderson et al. (2020) generated a hybrid data-driven approach by using multiple technologies to assess learners more accurately. In our prior studies, we proposed approaches to clustering learners to identify and visualize the in-game tools using patterns in an educational game we developed (Feng & Yamada, 2019; Feng & Yamada, 2020). Alonso-Fernandez et al. (2019) indicated that under game-based assessment research, an analytics based on new perspectives and a focus on more complex contextual data is needed for assessments to effectively evaluate learners' performances and provide evidence to improve the quality of educational games.

1.2. Learning paths and precision education

"Precision education" is a relatively new concept and is considered a challenging subject in the field of LA, especially with regard to artificial intelligence (AI) technology (Yang, 2019). The concept of precision education is based on personalized education. Its rationale is similar to precision medicine, which is meant to improve the precision, accuracy, and quality of treatment and disease prevention through an analysis of individual factors. Therefore, the goal of precision education is to apply LA technologies to individual learning factors, such as learning patterns or strategies, to identify and evaluate learning activities and effects, including identifying students who may be at risk of developing a learning disability, so as to provide them timely personalized educational intervention (Hart, 2016). For example, Tsai et al. (2020) used a statistical and an AI method (deep neural networks) to estimate students' probability of dropping out of university, and this approach was employed to identify the students expected to drop out and need timely interventions. Additionally, research examining possible influential factors can contribute to precision education. For example, Lu et al. (2018) examined 21 variables in a blended learning environment, and subsequently identified seven factors that can predict students' performance when only one-third of the semester has elapsed. This new concept has the potential to provide a new perspective for analytics in game-embedded assessment. Specifically, games are often open-ended, so players have personalized learning experiences. This feature shows the feasibility and suitability of integrating game-embedded assessment with precision education.

As for the objective being analyzed, we believe that the learning path is a piece of information that can be analyzed to create a personalized precision education experience. The learning path term generally refers to the route from preconception to the target, and usually includes the order in which the content is learned. Learning paths are considered one of the individual attributes of learning behavior and may influence learning effects (Williams & Rosenbaum, 2004; Shou et al., 2020). In an open-ended learning environment such as games or the e-learning system, the control over interactions within the learning environment shifts from the educator to the learner, so the learning paths of each individual learner vary. Accordingly, learning paths are a factor that might be analyzed in precision education. There are research studies on learning paths focusing on developing individual optimal learning path predictions and recommendation systems (Su, 2017; Shou et al., 2020), which seem to align with the precision education's goal, although there is no mention of precision education. For the game-embedded assessment, the learning path with regard to content is also a new data type that should be analyzed. As mentioned earlier, analysis focusing on context information is imperative, and the learning material related to in-game interaction is one form of context data constituting the learning path as well. Since research on learning content arrangement and instructional presentation is rare, despite Ruipérez-Valiente et al. (2019) proposing a visualization approach showing the path of completing tasks in the game, it is considered that the results can only provide insight into students' behaviors, and the teachers cannot assess the performance directly.

Based on the aforementioned discussion, we believe that exploring an analytical approach to identify learning path patterns and examining the relationship between them and the learning effects is expected to offset the limitations in game-based assessment research and aid precision education. Additionally, given that an understanding of the analysis results, for example explaining the identified patterns, is very important for understanding students' behaviors and improving precision education, an explanatory approach is also needed (Ke, 2016). Three research questions have been proposed below:

- Can the learning path detection approach we propose determine learning path patterns effectively?
- Do learners with diverse detected learning path patterns have different learning results? The answer will reveal whether the learning path can be an influential factor of learning and serve as an important variable in the precision education system.
- Can the explanatory approach proposed in this paper describe the identified pattern functionally to help us understand learners' behavior and promote the accuracy of precision education?

2. The educational game *Hist Maker*

2.1. General introduction

Hist Maker is an educational game developed by us for use in our research. The platforms for the game are PCs with a Windows OS and smartphones with an Android OS. Players can download the game online and play it where and when they wish, allowing them to learn in informal situations. It is a puzzle game intended to help players learn historical concepts related to world civilizations while solving a puzzle. The interfaces are shown in Figure 1.

Figure 1. Initial interface and gameplay interface of *Hist Maker*

There are three major components in the game:

- Core gameplay based on concept maps. Concept maps are considered effective cognitive tools (Novak & Cañas, 2008). The puzzles in the game were designed as concept maps of historical knowledge, which learners can explore while interacting with the game. The gameplay reflects the game characteristics of having a rich interactivity, which is considered to make the learning environment effective and motivational according to game-based learning theories (Shute & Ke, 2012). Also, this game is expected to impart knowledge about not only history, but also other subjects, such as biology and chemistry, that can be presented in the form of concept maps. The potential of generalization is a strength of Hist Maker, and the analytical approaches in our studies could have a broader scope. The core gameplay is closely related to the learning path of content acquisition, which we will discuss in the following sections.
- In-game tools. Based on game-based learning theories about how to make an educational game effective, some support tools were developed for the game. These tools include a task system, a hint tool, and a knowledge repository tool. The task system can provide clear goals to players and make the game an effective goal-oriented learning environment. The hint tool can lower the difficulty of tasks when players find it challenging, in order to provide challenges at the proper level (Gee, 2003), and the knowledge repository tool can work as a cognitive tool to reduce the memory load required when playing games (Bera & Liu, 2006).
- Data collection system. We developed a game-embedded system to help us collect data. An introduction to it along with its details is included in the section “4.3. Data Collection.”

2.2. Core gameplay based on concept maps

The core gameplay of Hist Maker is based on concept maps, which are considered an effective cognitive tool for imparting knowledge (Novak & Cañas, 2008). Concept maps are not shown in the game because Charsky and Ressler (2011) pointed out that showing a complex concept map at the beginning of the learning process may diminish learning motivation. The knowledge in concept maps is presented as “elements and formulas” in the puzzle. Each “element” represents a piece of knowledge or concept from history. When learners start a level in the game, there are only a few elements, but they can add new ones by combining two elements that have already been acquired. This format can be described using the formula: “element A + element B = element C.” The relationships in concept maps are also represented in this way, and the instructional text is shown when a formula is revealed. For example, for the formula “east of Yellow River Basin + Tribe = Dongyi Tribes,” the instructional text is “The Dongyi Tribes were the tribes located east of the Yellow River Basin.” Figures 2 and 3 show the correspondence between the concept map and the formulas.

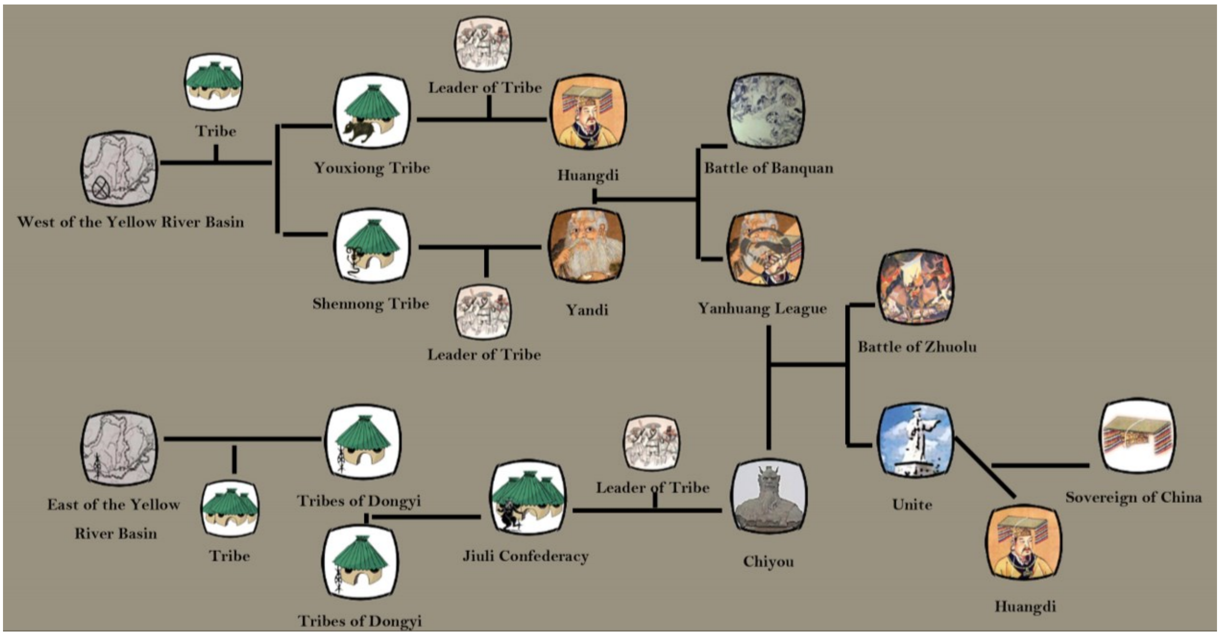


Figure 2. Part of a concept map in the level “The Five Sovereigns era”

In this way, players can explore the content in concept maps using multiple learning paths that involve combining the elements in various orders. Further, since concept maps emphasize the relationship between prior and new knowledge, the order of content acquisition from concept maps may be especially influential on learning effects (Chen, 2009). Therefore, detecting learning path patterns and modifying the design of the formula to eliminate the patterns that lead to negative learning effects is important.

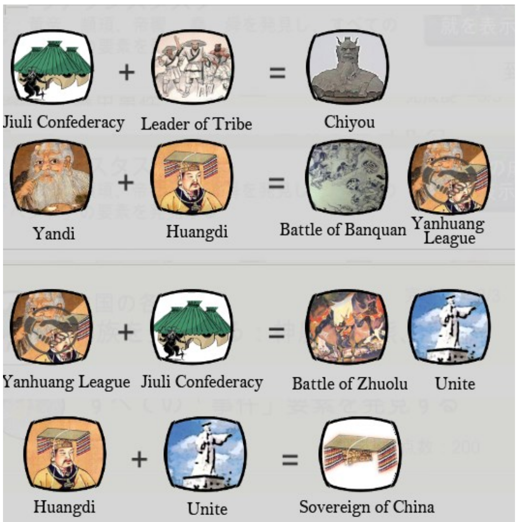


Figure 3. Examples of some of the formulas from the concept map

3. Analytical approach

To detect learning path patterns, the sequential characteristics of learners' content acquisition routes should be described objectively, and a method for grouping learning paths according to their common features is needed. In this study, we used the Levenshtein distance to demonstrate the sequential similarities between routes and then classify the learning sequences using hierarchical cluster analysis. The proposed analytical approach combines the Levenshtein distance and hierarchical cluster analysis.

3.1. The Levenshtein distance

The Levenshtein distance is a measure of distance proposed by the Soviet mathematician Vladimir Levenshtein to show the similarities between two character-strings, such that the higher the degree of similarity between two strings, the shorter the Levenshtein distance. This term refers to the minimum number of editing steps required to transform one character-string into another; thus, it is also termed the "edit distance." Edit operations include inserting, deleting, or replacing a character (Levenshtein, 1966).

The Levenshtein distance is widely used to demonstrate the degree of difference between words, sentences, or essays (e.g., Schepens, 2012). Moreover, the Levenshtein distance can be used with other types of sequential data, including learning paths. For example, Hao et al. (2015) used the Levenshtein distance to propose an approach to analyzing process data from game-based learning tasks. The idea behind the approach is to code each type of action into a single character so that the action sequences in the learning process can be transformed into character-strings. The Levenshtein distance can then be used to represent the distance between the learning sequences of players. In this vein, the Levenshtein distance can also be used to demonstrate the sequential character of the path of content acquisition. However, the approach of Hao et al. (2015) is only available for a learning environment that includes the best action sequence. Additionally, sequences of learning content have not been considered in their study. Our research proposes an original method based on Hao et al. (2015)'s approach to detect learning content path patterns.

3.2. Hierarchical cluster analysis

Cluster analysis is a proper means of detecting "patterns" in a situation in which a clear classification criterion is lacking. The procedure of cluster analysis involves grouping samples based on the distances between them. Samples close to each other are automatically classified as a group (cluster). Samples in the same cluster share common features because short distances represent similarity; these common features can be regarded as the profile of a “pattern.” Therefore, when distances indicate the characteristics of players’ behaviours and routes of content acquisition, cluster analysis can be used to detect learning path patterns.

To detect learning path patterns, we used hierarchical cluster analysis, which is a way to construct clusters in a hierarchical order. The algorithm is as follows: each sample belongs to a cluster that contains only itself at the beginning; then, the pairs of clusters with the shortest distances are merged into higher-level clusters, and this merging operation is repeated until a specified number of higher-level clusters remains. The merits of hierarchical cluster analysis make it especially well-suited for our research. According to Alonso-Fernandez’s (2019) review, clustering is a relatively popular approach for game-based learning research utilizing data science. Although there are various types of cluster analysis, we chose the hierarchical cluster analysis on account of two main merits: One is that the process of clustering can be visualized in the form of a dendrogram, and the other is that it has wide applicability and can be applied to various types of distances. The first merit can help us to determine the final number of clusters based on the resulting merging dendrogram. The second merit is that the algorithm is guaranteed be suitable for dealing with the Levenshtein distance. An example of the application of hierarchical cluster analysis to character-strings is shown in Figure 4.

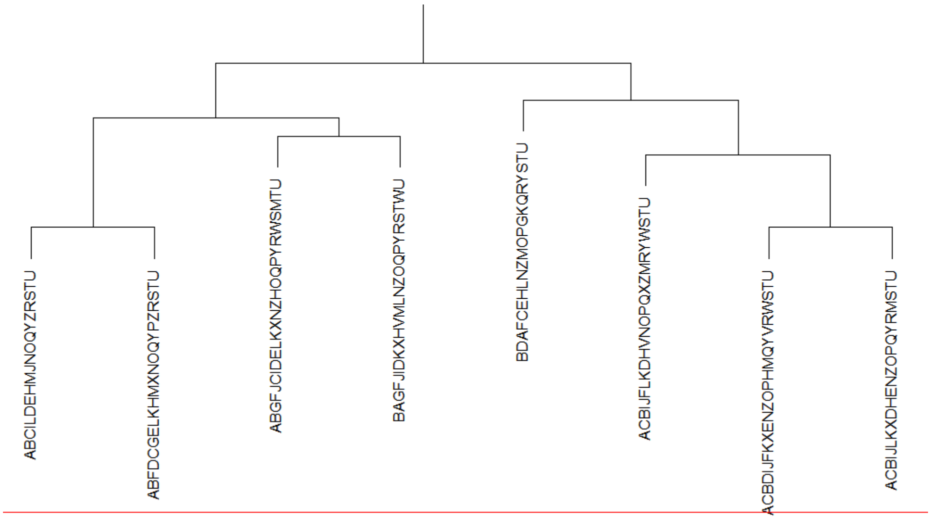


Figure 4. Example of a merging dendrogram of hierarchical cluster analysis

3.3. Analytical approach of this study

Based on the definitions of the terms “Levenshtein distance” and “hierarchical cluster analysis,” the analytical approach we propose in this study will be illustrated in this section.

As mentioned above, learners primarily acquire knowledge from instructional texts when attempting to identify the “formulas” in a level. Therefore, the order of actions taken to identify formulas while completing a level is considered the learning path of content, and the learning pattern can be detected from an analysis of the data. Concretely, the analytical approach has four steps:

- Step 1, coding the data. This step is based on the idea that various types of sequential data can be converted into character-strings by coding each piece of single-item data as a unique character. Thus, the first step is to code each formula as a letter of the alphabet so that the learning acquisition path of each learner can be transformed into a character-string.
- Step 2, calculating the Levenshtein distance. Since character-strings can be obtained through Step 1, the Levenshtein distance can be calculated, and the distance can be used to represent the sequential characteristics and the similarities between the learning paths of learners.
- Step 3, implementing a hierarchical cluster. Based on the calculated Levenshtein distance, a hierarchical cluster analysis can be used to classify learners into several clusters. Learners in the same cluster demonstrate common features in their learning path patterns, meaning that each cluster represents a pattern.
- Step 4, explaining the results. To draw educationally meaningful conclusions, the relationships between these patterns and learning effects or gameplay times, as well as the specific features of each cluster, should be discussed. Therefore, the correlations should be examined using statistical methods, and the pattern of each cluster must be presented and explained.

4. Methods

4.1. Participants

To guarantee the randomness of the informal learning behaviors, players were not intentionally recruited. The players played the game and submitted the data completely spontaneously. The data collection was limited to the level “Chinese history: The Five Sovereigns era” in the game, and the participants were all Chinese.

Consequently, we collected valid data from 548 players, and the distribution patterns with regard to their gender and educational status are shown in Figure 5: 423 were male, 110 were female, and 15 failed to indicate their gender. Regarding the educational background, 127 were primary school students, 143 were middle school, 85 were high school students, 110 were university or graduate students, 66 were “other,” including those who worked, and 17 players failed to specify their educational status. These proportions seemed unbalanced, especially for gender, but it is also proof that players chose the game spontaneously. Moreover, the average age is 16.73 (*SD* = 5.38), but 58 players did not answer this question or gave an apparently false answer such as 100 years old.

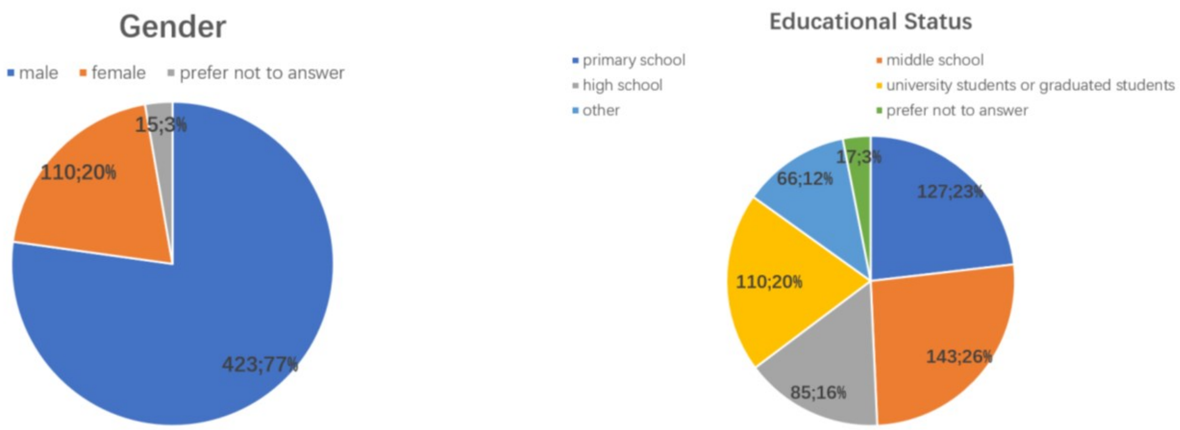


Figure 5. Distribution patterns of learners’ gender and educational status

4.2. Learning game content

In this study, we limited our data collection and analysis to the level “Chinese history: The Five Sovereigns era.” The content of this stage pertained to the “Legendary Era” of prehistoric China before the Xia Dynasty. The reason we chose this era is that our plan of setting the game stages is to show the development of civilizations in chronological order, and the Five Sovereigns era is the first era in Chinese history. Although it is a prehistorical era without adequate archaeological evidence, this period is considered historical because of literary evidence and is included as content in formal history textbooks. Analyzing students’ behaviors in this stage may help us design the stages about follow-up eras with better quality. The content presented in the level included a historical story from the *Records of the Grand Historian* (*Shiji*) of Sima Qian, which is considered the most well-known piece of historical literature about ancient China and is highly authoritative. To provide a flexible, informal learning environment, we decided not to include a time limit for game play, and the players did not need to explore all the formulas. Clearing a level indicated that most of the important knowledge in it had been learned before taking the post-test.

Specifically, the content of the levels was divided into four categories: The main and secondary parts of the pre-unification era, and the main and the secondary parts of the post-unification era. The boundary between the pre- and the post-unification eras is taken as the event when Huangdi defeated the other tribes and unified China area. The main part refers to the knowledge of historical figures and important events in the era, while the secondary parts mostly included information about the activities of historical figures and their contribution to society. Since the analysis required the formulae from this stage to be coded into alphabetical characteristics, a part of the correspondence table is shown in Table 1.

Table 1. Part of the correspondence table of formulae and alphabetical codes

Formula	Alphabetical Code
Tribe+Yellow River Basin West = Tribe Youxiong+Tribe Shennong	A
Tribe+Yellow River Basin East = Tribes in Dongyi	B
Tribe Youxiong+Tribe Shennong = Battle of Banquan+Confederacy of tribes Yanhuang	C
…	…
Nature+Yandi = Agriculture	K
Nature+Huangdi = Calendar	L
Nature+Chiyou = Bronze weapon	M
Huangdi+Unifying China = Sovereign of China	N
Son generation+Huangdi = Shaohao	O
Shaohao+Tribes in Dongyi = Feng Totem	P
Shaohao+Son generation = Zhuanxu	Q
Zhuanxu+Son generation = Diku	R
Diku+Son generation = Yao	S
Yao+Sovereign of China = Demising	T
Yao+Demising = Shun	U
…	…
Huangdi+Sovereign of China = Long Totem	Z

4.3. Data collection

The data collection system embedded in the game was developed for both pre/post-test and game-embedded assessment. The database structure of the game is shown in Figure 6. The players had to complete the pre-test before beginning the game; subsequently, the operation log data were recorded during gameplay, and in the end, when players cleared a level, they completed the post-test and sent us the data on the test answers and operation log. We did not utilize database software such as MySQL, and the collected data are saved in the local storage of the device in *.csv format and sent to us by email.

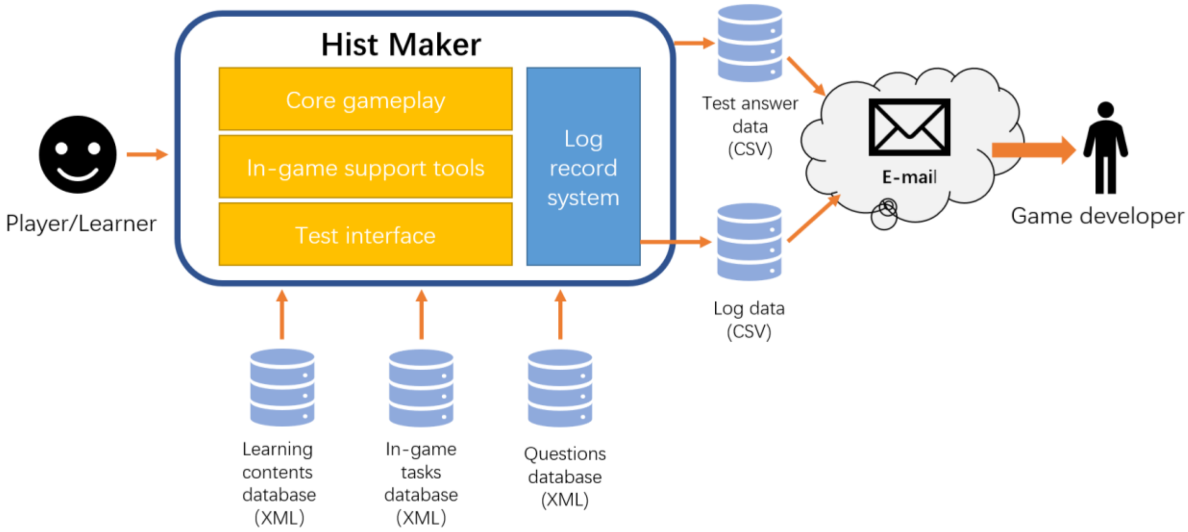


Figure 6. The database structure of the game

The test for each stage included seven multiple-choice questions. These questions are prepared based on the learning content in the game, and following discussions with a history teacher in a high school in mainland China, we modified the questions to ensure reliability and validity. Each question had three answer options and an “I don’t know” option, and there was only one correct response for each question. Therefore, the test is presented as follows: “Question: The Yanhuang Alliance fought in the land of () and defeated Chiyou, an important event leading to the unification of the Chinese tribe. Answer: A. Zhuolu; B. Banquan; C. Jizhou; D. I don’t know.” An analysis of the test results can directly show the learning effects on learners’ mastery of the content and how their level of knowledge changed while playing the game. Providing correct answers to all the questions indicated the best learning effect.

The recorded gameplay log data included not only the type and frequency of each operation but also various pieces of contextual information, such as timestamps and information about the elements and formulas that learners identified. Some examples of recorded log data are shown in Table 2. Therefore, the sequence of actions taken to identify formulas can be extracted from the log data. Further, we also collected some pieces of demographic information, such as the gender and educational background of each learner, through pre-questionnaires.

Table 2. Examples of recorded log data

Player ID	Type of operation	Contextual information 1	Contextual information 2	Contextual information 3	Contextual information 4
1001	Select an element	4/24/2019 9:00:36	Right	ELE_HUANGDI	CATE_FIGURE
1005	Attempt to combine elements	5/12/2019 15:28:47	ELE_HUANGDI	ELE_TRIBE	Fail
1020	Close the instruction text panel	3/4/2019 20:01:55	LEADER_PLUS_TRIBE_YOUXIONG	Null	Null
1035	Request a hint	6/22/2019 13:15:00	Hint tool	Hint exists	Null

5. Results

5.1. Detecting learning path patterns and examining learning effects

We implement the proposed approach by programming in R language. As a result of the utilization of the analytical approach, we obtained three clusters of learners, each believed to represent a specific learning path pattern. Specifically, 166 learners were in Cluster 1; 374 learners were in Cluster 2; and only 8 learners were in Cluster 3. Although the results were unbalanced, the clusters were determined to be far enough apart based on the merging dendrogram of the hierarchical cluster analysis. Cluster 3 was much smaller than the others; thus, it may be worthwhile discussing the learning pattern in this cluster.

In this study, the results of the pre- and post-tests demonstrate learning effects. However, each test question included the option "I don't know," which cannot be considered a wrong answer; thus, a simple calculation of the test scores is not suitable for measurement. To address this problem, we coded the changes in answers to the same questions on the pre- and post-tests with reference to White's research (2012) that dealt with a similar situation. The criteria of the code are shown in Table 3, and the information about the players' test answers and length of gameplay is shown in Table 4.

Table 3. Criteria of coding

Answer in Pre-test	Answer in Post-test	Code
Wrong answer or "I don't know"	Correct answer	"Better"
Correct answer	Correct answer	"Equivalent"
Correct answer	Wrong answer or "I don't know"	"Worse"
"I don't know"	Wrong answer	"Misunderstood"
"I don't know"	"I don't know"	"No effect"
Wrong answer	Wrong answer or "I don't know"	"No effect"

Table 4. Results of all players

Attribute	Mean	*SD*	Maximum	Minimum
Number of Code "Better"	2.16	1.47	7	0
Number of Code "Equivalent"	2.28	1.82	7	0
Number of Code "Worse"	0.39	0.64	4	0
Number of Code "Misunderstood"	0.62	1.13	7	0
Number of Code "No effect"	1.55	1.45	7	0
Length of gameplay (milliseconds)	1788313	5009032	93585705	451200

The number of each type of code can demonstrate the overall circumstances of learning effects for a single cluster or individual learner. For example, the number of "Better" responses can show how players acquired new knowledge with a lack of prior knowledge or how they corrected misconceptions; the number of “Equivalent” responses is also related to a relatively positive learning effect, since it shows instances in which correct knowledge of the concept was maintained.

Therefore, to examine whether learning path patterns can influence learning effects, we used one-way analysis of variance and post-hoc comparisons to identify the differences between the clusters’ means for each type of code. The results of the statistical analysis are shown in Table 5. Based on these results, we observe that there are significant differences between the clusters in the codes “Equivalent,” “Misunderstood,” and “No effect.” The three clusters did not demonstrate any significant differences in length of gameplay.

Table 5. Results of one-way analysis of variance and post-hoc comparisons

Attribute	Cluster 1 *M* (*SD*)	Cluster 2 *M* (*SD*)	Cluster 3 *M* (*SD*)	*F*	*p*	Post-hoc Comparison
Number of Code “Better”	2.25 (1.45)	2.13 (1.48)	1.50 (1.20)	1.29	.28	None
Number of Code “Equivalent”	2.54 (1.76)	2.20 (1.84)	0.63 (0.92)	5.50	$.004^*$	$C1>C3^*$; $C2>C3^*$
Number of Code “Worse”	0.39 (0.61)	0.40 (0.66)	0.50 (0.76)	0.12	.88	None
Number of Code “Misunderstood”	0.40 (0.75)	0.72 (1.25)	0.75 (1.04)	4.52	$.01^*$	$C2>C1^*$
Number of Code “No effect”	1.41 (1.14)	1.56 (1.54)	3.63(1.41)	9.293	$< .001^*$	$C3>C1^*$ $C3>C2^*$
Length of gameplay (milliseconds)	1677317 (2903344)	1851062 (5748510)	1157924 (373811)	0.13	.88	None

Note. $^*p < .05$.

5.2. The relationship between learning path patterns and learning effects

The mean value for a code can indicate the situation of the learning effects of a cluster; for the codes that are significantly different, it is believed that the code “Equivalent” indicates relatively positive effects of keeping correct prior knowledges, while “Misunderstood” and “No effect” suggest negative effects. Accordingly, based on a comparison of the specific significant differences between clusters, we concluded that players in different clusters experienced diverse effects, which also implies correlations between learning path patterns and learning effects. Based on the results of the post-hoc comparison, it can be concluded that the learning effects of the players in Cluster 1 were more positive than those of their counterparts in Cluster 2; as for the learning effects of the players in Cluster 3, they represented a distinctive learning pattern used by the few players who fared the worst.

5.3. Explanation of learning path patterns

Besides discussing the relationships between learning path patterns and learning effects, it is important to explain the concrete patterns of learning paths. Even when traditional statistical methods can be used to confirm that different patterns of content acquisition lead to different effects, the results still do not tell us what the learning path patterns are for clusters. Without this information, modifying the design of a game to eliminate a “bad learning path” and improve the game’s quality remains difficult. To elucidate the learning path patterns, especially in the clusters that included many learners, systematic methods of analysis need to be used. In this study, we utilized two methodologies: sorting out the most common sub-sequences and building a process to extract a representative sequence from each cluster.

5.3.1. Finding the most common sub-sequences

Since the learning path pattern-detection approach is based on the Levenshtein distance and hierarchy cluster analysis, it is easy to conclude that there should be short Levenshtein distances between learners in the same cluster. Applying the definition of Levenshtein distance, we inferred that characteristic-strings with short Levenshtein distances shared more common sub-sequences. The length of valuable common sub-sequences may not be great for a large cluster because the proportion of long sub-sequences is often very small, and the representation is not adequate. Therefore, the proportions and quantity are assigned more importance than the length of the sub-sequences.

The learning path characteristic-strings and the results of efforts to detect the most common sub-sequences of the eight learners in Cluster 3 are shown in Tables 6 and 7. Based on the results, we conclude that the sub-sequence "STU" made up a very large proportion and that this sub-sequence was present at the end of the learning acquisition sequence. "STU" represents formulas related to the last two sovereigns in the level "The Five Sovereigns era": Yao and Shun. Thus, the learning path for Cluster 3 included learning all the stories about successions to the throne and the contributions of other sovereigns, then gaining knowledge about Yao and Shun.

Table 6. Learning path characteristic-strings in Cluster 3

Temporary Player ID	Learning path
3-1	ACBDIJFKXENZOPHMQYVRWSTU
3-2	ABGFJCIDELKXNZHOQPYRWSMTU
3-3	BAGFJIDKXHVMLNZOQPYRSTWU
3-4	ACBIJFLKDHVNOPQXZMRYWSTU
3-5	ACBIJLKXDHENZOPQYRMSTU
3-6	ABCILDEHMJNOQYZRSTU
3-7	ABFDCGELKHMXNOQYPZRSTU
3-8	BDAFCEHLNZMOPGKQRYSTU

Table 7. Parts of common sub-sequences in Cluster 3

Sub-sequence	Frequency
ST	7
TU	7
STU	6
NZ	5
KX	4
OP	4
QY	4
LK	4
OQ	4

Table 8. Parts of the most common sub-sequences and comparison between Clusters 1 and 2

Sub-sequence	Frequency in Cluster 1	Proportion in Cluster 1	Frequency in Cluster 2	Proportion in Cluster 2	Difference Between Proportions (Cluster 1− Cluster 2)	Difference Between Proportions (Absolute Value)
NOQRS	134	80.72%	225	60.16%	20.56%	20.56%
NOQR	135	81.33%	230	61.50%	19.83%	19.83%
NOQ	146	87.95%	257	68.72%	19.24%	19.24%
MN	2	1.20%	72	19.25%	−18.05%	18.05%
LM	50	30.12%	47	12.57%	17.55%	17.55%
MNO	2	1.20%	70	18.72%	−17.51%	17.51%
MNOQ	1	0.60%	65	17.38%	−16.78%	16.78%
NO	153	92.17%	284	75.94%	16.23%	16.23%
SW	3	1.81%	67	17.91%	−16.11%	16.11%
RSK	28	16.87%	3	0.80%	16.07%	16.07%
SK	28	16.87%	3	0.80%	16.07%	16.07%
HM	4	2.41%	69	18.45%	−16.04%	16.04%
RSW	3	1.81%	66	17.65%	−15.84%	15.84%
NOQRSL	27	16.27%	2	0.53%	15.73%	15.73%
MNOQR	1	0.60%	61	16.31%	−15.71%	15.71%
OQRSL	27	16.27%	3	0.80%	15.46%	15.46%
QRSL	27	16.27%	3	0.80%	15.46%	15.46%
RSL	27	16.27%	3	0.80%	15.46%	15.46%
SL	27	16.27%	3	0.80%	15.46%	15.46%
MNOQRS	1	0.60%	59	15.78%	−15.17%	15.17%

As for the large clusters, Clusters 1 and 2, existing featured sub-sequences accounting for large proportions were unusual, so drawing a comparison between the proportions of sub-sequences enabled us to discuss the specific characteristics of the learning path patterns. Thus, we arranged the sequences in order of the differences in proportions between Clusters 1 and 2. The results of our effort to detect the most common sub-sequences and a comparison of them are shown in Table 8; the results were sorted by the absolute values of the differences. From these results, we observe that the proportions of the sequences "MNOQRS," "RSW," and some of their sub-sequences were relatively high in Cluster 2, while the proportions of these sequences in Cluster 1 were extremely low. Contrariwise, the learning path sequences in Cluster 1 included many instances of "NOQRSL," "RSK," and some of their sub-sequences, while Cluster 2 showed few examples of these patterns. Further, although "NOQRS" and some of its sub-sequences accounted for the greatest differences between the two clusters, they still accounted for a relatively large proportion of both clusters.

"NOQRS" included the formulas related to Huangdi becoming the sovereign of China and hereditary succession until the reign of Sovereign Yao. Since in all these successions, the sovereigns passed the throne to a family member of the next generation, such as a son or nephew, players easily discovered that the formula "one of the sovereigns + Son generation" would work in several instances; thus, "NOQRS" accounted for a large proportion of both clusters. The alphabetical codes "K," "L," and "M" represented the formulas related to the contributions of Huangdi, Yandi, and Chiyou, three important historical figures in the pre-unification era. Thus, it can be observed that the learners in Cluster 1 tended to learn about the secondary part of the "pre-unification era" section after learning about the primary part of the "post-unification era" section, while the order of content acquisition was reversed in Cluster 2.

6. Discussion

Using these two methods, we can conclude that the presence of the "STU" learning formula at the end of the sequence was the distinctive feature of Cluster 3's learning path pattern. The difference between Clusters 1 and 2 can be seen in the order in which the participants learned about the secondary part of the pre-unification era section and the primary part of the section about the post-unification era.

We inferred that learners in Cluster 3 experienced negative learning effects perhaps because the story of Yao and Shun is familiar to most Chinese, so the content covered in the "STU" pattern could be considered prior knowledge to some degree. Information about other sovereigns, especially those who ruled after Huangdi, is not quite as widely known, but these sovereigns had some relationship with Yao and Shun. Thus, when comparing the clusters, it appears that the learners in Cluster 3 built fewer connections between new knowledge about other sovereigns and their prior knowledge about Yao and Shun because they did not learn anything about Yao or Shun until the end of the level. Although the sample size of Cluster 3 was so small so that the persuasiveness of the inference was compromised, this inference can still provide a possible direction for modifying the game's design.

Regarding why the players in Cluster 1 experienced the most positive learning effects, based on the theory of the cognitive structures of narrative discourse (Thorndyke, 1977), we propose that the order in Cluster 1 indicates that there is a strong temporal and causal character involved in learning the primary parts of the content related to the pre-unification era and the post-unification era; thus, the learning paths in Cluster 1 are constructed relatively efficiently, which may lead to especially positive learning effects.

There was no significant difference between the clusters in terms of the length of gameplay, which means that although the amount of time spent learning can be an influential factor in constructing learning paths, it was not an important variable in this study. We found that the standard deviation of time spent playing the game was extremely large; we believe that this may be because the length of gameplay was affected by many factors besides behavioral patterns and content acquisition routes and that there may have also been some factors in the external environments of the players that were difficult to infer from the log data alone.

Discussion of the reasons that different patterns may lead to different effects can help us obtain clues to improving the quality of educational games. Since we can use the rules guiding combinations of elements to control the order in which players discover formulas and learn pieces of knowledge, we can modify the design to help players avoid taking a "bad learning path" and possibly help them choose a more effective path instead.

7. Conclusion

Our studies revolved around the development of analytical approaches by applying LA technology to build a game-embedded assessment to examine the potential value of learning paths in precision education for improving the quality of the educational game. The approach is centered on cluster analysis, which is considered a type of AI method. We applied the analytical approach to the educational game *Hist Maker*, and 548 samples were categorized into three clusters, with each representing a learning path pattern. Specifically, a minor pattern has been detected and learners with this pattern have significantly worse learning effects than others; hence, this pattern can indicate an "at risk" status for precision education. The results indicated that the proposed approach can detect non-normative (but featured) patterns, which we consider a meaningful tool for supporting the realization of precision education. After the statistical examination and discussion of the learning patterns in each cluster, we conclude that the consideration of learning paths can help us to realize precision education in a game-based learning environment. Additionally, we proposed a method to explain the specific learning path patterns in each cluster. The practice in this study proved that explanatory methods can help developers understand the analytical results and consider precision education from a game design perspective.

7.1. Contributions of this study and implications for practice

This study provides empirical evidence that the learning path with regard to learning content can influence learning in a game-based environment. For the precision education system that regards "having a bad learning effect" as "risk," the results of the analysis indicate that the learning path is a factor needing consideration. We encourage other researchers to focus on contextual data of interactions that are directly related to content, such as learning paths, in order to improve the accuracy of precision education predictions. Additionally, this study proposed a practically validated analytical approach to detect and interpret learning path patterns in *Hist Maker*. Nevertheless, while changing the granularity of the coded content, this approach also has potential for application to other learning environments based on concept maps and even all learning environments that aim for knowledge acquisition. Lastly, the proposed approach also includes a method of interpreting detected patterns. With the discussion of specific contents and cognitive psychology theories, this approach offers a comprehensive understanding and a deeper insight into learners' behavior patterns and cognitive processes. We encourage other researchers to explore approaches that consider the results of such analysis to gain a deeper understanding of learner behavior in order to provide more evidence for designing a precision education system.

7.2. Recommendations and future works

Despite the high suitability of the analytical approach in this study, as identified in practice, some work remains for the future. First, it is imperative to broaden the examination of the applicability of the proposed approach beyond the single game level examined in this study. The versatility of the analytical approach needs to be examined for other learning content, other game levels, and even other learning environments. Second, this study concluded that the learning path is an important influential factor in precision education, although variables such as the learner's personal attributes and the device platform can influence the learning effects and learning behaviors, and the optimal learning path for each learner is also governed by these variables (Su, 2017). However, they were not collected or analyzed in this study. Therefore, we recommend that future research examine the interaction between learning paths and these variables and the overall effect on learning effects in order to realize personalized interventions in precision education. Third, this study proposed an approach to analyze already collected data, but a precision education system still needs algorithms that can predict learning effects through an analysis of learning paths in the process (Hart, 2016). It will be necessary in the future to explore an approach that presents the right time to provide interventions.

Acknowledgement

This research is supported by JSPS KAKENHI Grants Nos. JP19H01716 and JP18K18657.

References

Akram, B., Min, W., Wiebe, E., Mott, B., Boyer, K. E., & Lester, J. (2018). Improving stealth assessment in game-based learning with LSTM-based analytics. In *Proceeding of the 11th International Conference on Educational Data Mining* (pp. 208-218). Raleigh, NC: International Educational Data Mining Society. Retrieved from https://files.eric.ed.gov/fulltext/ED593099.pdf

Alonso-Fernandez, C., Calvo-Morata, A., Freire, M., Martinez-Ortiz, I., & Fernández-Manjón, B. (2019). Applications of data science to game learning analytics data: A Systematic literature review. *Computers & Education*, *141*, 103612. doi:10.1016/j.compedu.2019.103612.

Bera, S., & Liu, M. (2006). Cognitive tools, individual differences, and group processing as mediating factors in a hypermedia environment. *Computers in Human Behavior*, *22*(2), 295–319. doi:10.1016/j.chb.2004.05.001.

Charsky, D., & Ressler, W. (2011). "Games are made for fun": Lessons on the effects of concept maps in the classroom use of computer games. *Computers & Education*, *56*(3), 604–615.

Chen, C. M. (2009). Ontology-based concept map for planning a personalised learning path. *British Journal of Educational Technology*, *40*(6), 1028–1058.

Feng, X., & Yamada, M. (2019). Effects of game-based learning on informal historical learning: A Learning analytics approach. In *Proceedings of the 2019 27th International Conference on Computers in Education* (pp. 505-514). Kenting, Taiwan: Asia-Pacific Society for Computers in Education.

Feng, X., & Yamada, M. (2020). Design and development of visualization approaches for informal learning game logs. In *Proceedings of the 2020 DiGRA International Conference: Play Everywhere*. DiGRA. Retrieved from http://www.digra.org/wp-content/uploads/digital-library/DiGRA_2020_paper_104.pdf

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE), 1*(1), 1–4.

Hao, J., Shu, Z., & von Davier, A. (2015). Analyzing process data from game/scenario-based tasks: An Edit distance approach. *Journal of Educational Data Mining*, *7*(1), 33–50.

Hart, S. A. (2016). Precision education initiative: Moving toward personalized education. *Mind, Brain, and Education*, *10*(4), 209−211.

Henderson, N., Kumaran, V., Min, W., Mott, B., Wu, Z., Boulden, D., Lord, T., Reichsman, F., Dorsey, C., Wiebe, E., & Lester, J. (2020). Enhancing student competency models for game-based learning with a hybrid stealth assessment framework. In *Proceedings of the 13th International Conference on Educational Data Mining (EDM 2020)* (pp. 92–103). Online: International Educational Data Mining Society. Retrieved from https://files.eric.ed.gov/fulltext/ED607823.pdf

Ifenthaler, D. (2015). Learning analytics. In J. M. Spector (Ed.), *The SAGE encyclopedia of educational technology* (Vol. 2, pp. 447−451). Thousand Oaks, CA: Sage.

Ke, F. (2016). Designing and integrating purposeful learning in game play: A Systematic review. *Educational Technology Research & Development*, *64*, 219–244. doi:10.1007/s11423-015-9418-1

Kim, Y. J., & Ifenthaler, D. (2019) Game-based assessment: The Past ten years and moving forward. In D. Ifenthaler, & Y. Kim (Eds), *Game-Based Assessment Revisited. Advances in Game-Based Learning* (pp. 3-11). Cham, Switzerland: Springer. doi:10.1007/978-3-030-15569-8_1

Klein, J. D., & Freitag, E. (1991). Effects of using an instructional game on motivation and performance. *Journal of Educational Research*, *84*(5), 305–311.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady, 10*(8), 707–710.

Loh, C. S. (2011). Using in situ data collection to improve the impact and return of investment of game-based learning. In *Proceedings of ICEM-SIIE 2011, the 61st International Council for Educational Media (ICEM) and the XIII International Symposium on Computers in Education (SIIE) Joint Conference* (pp. 801–811). Aveiro, Portugal: ICEM-SIIE.

Lu, O. H., Huang, A. Y., Huang, J. C., Lin, A. J., Ogata, H., & Yang, S. J. (2018). Applying learning analytics for the early prediction of students' academic performance in blended learning. *Educational Technology & Society*, *21*(2), 220-232.

Novak, J. D., & Cañas, A. J. (2008). The Theory underlying concept maps and how to construct and use them. *Technical Report IHMC CmapTools 2006-01 Rev 01-2008*. Pensacola, FL: Florida Institute for Human and Machine Cognition.

Ruipérez-Valiente, J. A., Rosenheck, L., & Kim, Y. J. (2019) What does exploration look like? Painting a picture of learning pathways using learning analytics. In D. Ifenthaler, & Y. Kim (Eds), *Game Based Assessment Revisited* (pp. 281–300). Cham, Switzerland: Springer. doi:10.1007/978-3-030-15569-8_14

Schepens, J. (2012). Distributions of cognates in Europe as based on Levenshtein distance. *Bilingualism: Language and Cognition*, *15*(1), 157–166.

Shou, Z., Lu, X., Wu, Z., Yuan, H., Zhang, H., & Lai, J. (2020). On learning path planning algorithm based on collaborative analysis of learning behavior. *IEEE Access*, *8*, 119863-119879.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 503–524). Charlotte, NC: Information Age.

Shute, V. J., & Ke, F. (2012). Games, learning, and assessment. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning* (pp. 43–58). New York, NY: Springer.

Su, C. (2017). Designing and developing a novel hybrid adaptive learning path recommendation system (ALPRS) for gamification mathematics geometry course. *Eurasia Journal of Mathematics*, *Science and Technology Education*, *13*(6), 2275–2298.

Thorndyke, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, *9*(1), 77−110.

Tsai, S., Chen, C., Shiao, Y., Ciou, J., & Wu, T. (2020). Precision education with statistical learning and deep learning: A Case study in Taiwan. *International Journal of Educational Technology in Higher Education*, *17*, 1–13.

Werquin, P. (2007). Terms, concepts and models for analyzing the value of recognition programmes. In *Report to RNFIL: Third Meeting of National Representatives and International Organisations* (pp. 2-3). Vienna, Austria: OECD. Retrieved from http://www.oecd.org/dataoecd/33/58/41834711.pdf

White, B. J. (2012). A Conceptual approach to the instruction of phrasal verbs. *The Modern Language Journal*, *96*(3), 419−438.

Williams, J., & Rosenbaum, S. (2004). *Learning paths: Increase profits by reducing the time it takes employees to get up-to-speed*. Hoboken, NJ: John Wiley & Sons.

Yang, S. J. H. (2019). Precision education: New challenges for AI in education [conference keynote]. In *Proceedings of the 27th International Conference on Computers in Education (ICCE)* (pp. XXVII-XXVIII). Kenting, Taiwan: Asia-Pacific Society for Computers in Education (APSCE).

Han, F., & Ellis, R. A. (2021). Predicting Students' Academic Performance by Their Online Learning Patterns in a Blended Course: To What Extent Is a Theory-driven Approach and a Data-driven Approach Consistent? *Educational Technology & Society*, *24* (1), 191-204.

Predicting Students' Academic Performance by Their Online Learning Patterns in a Blended Course: To What Extent Is a Theory-driven Approach and a Data-driven Approach Consistent?

Feifei Han[1, 2*] and Robert A. Ellis[1]
[1]Office of Pro-Vice-Chancellor (Arts, Education and Law), Griffith University, Australia // [2]Griffith Institute for Educational Research, Griffith University, Australia // feifei.han@griffith.edu.au // r.ellis@griffith.edu.au
[*]Corresponding author

ABSTRACT: One of the major objectives of precision education is to improve prediction of educational outcome. This study combined theory-driven and data-driven approaches to address the limitations of current practice of predicting learning outcomes only using a single approach. The study identified the online learning patterns by using students' self-reported approaches and perceptions of online learning and by using the observational digital traces of the sequences of their online learning events in a blended course. The study examined predictions of the academic performance using the online learning patterns generated by the two approaches separately. It also investigated the extent to which the online learning patterns identified by the two approaches were associated with each other. The theory-driven approach adopted a hierarchical cluster analysis using the self-reported data and found a 'deep' and a 'surface' online learning patterns, which were related to differences in the academic performance. The data-driven approach used an agglomerative sequence clustering and detected four patterns of online learning, which not only differed by quantity (number of learning events), but also differed by quality (the proportions of types of learning events). A one-way ANOVA revealed that the online learning pattern which had the most learning events, and was characterized by high proportions of viewing course contents and of performing problem-solving exercises, had the highest academic performance. A cross-tabulation revealed significant association between the self-reported and observational online learning patterns, demonstrating consistency of the evidence by a theory-driven and a data-driven approach and triangulating the results of the two approaches.

Keywords: Online learning patterns, Academic performance, Theory-driven approaches, Data-driven approaches, Blended course designs

1. Introduction

Blended course design, which is "a systematic combination of co-present (face-to-face) interactions and technologically-mediated interactions between students, teachers and learning resources" (Bliuc, Goodyear, & Ellis, 2007, p. 234), has been increasingly adopted in the higher education sector worldwide. As blended courses require students to move back and forth across face-to-face and online contexts (Ellis & Goodyear, 2019), their learning experiences are related to an increasing number of core elements; their cognitions (e.g., conceptions, approaches, and perceptions, Trigwell & Prosser, 2020), their social interactions in learning (Hadwin & Oshige, 2011), the spaces in which they learn (Ellis & Goodyear, 2016), and the different devices they use for learning (Laurillard, 2013). As a result, student learning experiences are becoming more and more complex in their structure. Consequently it requires research methods that move beyond approaches that do not routinely investigate the combined contribution of participants and the things they use to outcomes, such as academic performance (López-Pérez, López-Pérez, & Rodríguez-Ariza, 2011; Wu, Tennyson, & Hsia, 2010). The recent precision education initiative has offered some alternative approaches to framing how we evaluate learning in universities (Hart, 2016, Williamson, 2019, Yang, 2019).

Precision education is based on the philosophy of the Precision Medicine Initiative launched by the former US present Barack Obama after his 2015 State of the Union Address (The White House, 2015). The Precision Medicine Initiative aimed to revolutionize the medical treatment of diseases by transiting away from the one-size-fits-all approach of medical research and practice to the personalized approach, which takes into account of individual differences in genetics, environments, and personal choices (Collins & Varmus, 2015). Rather than producing unique treatments for specific patients, precision medicine emphasizes individual variability to improve the diagnosis, prediction, treatment, and prevention of disease. Underpinned by the same principles, Lu et al. (2018) defines the objectives of precision education as "the improvement of diagnosis, prediction, treatment, and prevention of learning outcome" (p. 221). To achieve these aims, educational researchers have been increasingly using big data analytics, applying artificial intelligence and machine learning, and implementing advanced data mining techniques and complex algorithms (known as data-driven approaches) to identify at-risk students early, and to provide an prediction of students' academic performance by their learning behaviors; so that targeted intervention strategies can be planned in order to prevent drop-out and learning failure (Cook, Kilgus, & Burns, 2018; Tsai, Chen, Shiao, Ciou, & Wu, 2020). Another important focus of precision education lies in the personalized education for enhancing student learning experiences (Lemons et al., 2018; Rojas-López & García-Peñalvo, 2019; Wilson & Ismaili, 2019). To contribute to the development of how a precision education perspective can improve our understanding of university student experiences of learning, this research will focus on improving prediction of students' learning outcomes.

Traditionally, research into student learning experiences and academic performance in higher education has largely adopted theory-driven approaches, which test hypotheses derived from theories in educational psychology, learning sciences, and research in pedagogy and curriculum (Trigwell & Prosser, 2020). Studies using such approaches primarily employ self-reported instruments and data to understand the relations between the processes of learning (e.g., approaches to, and perceptions of, learning) and the product of learning (e.g., the academic performance) (Ellis & Goodyear, 2016). With the advancement of educational data mining techniques and collection of rich learning analytic data, data-driven approaches have gradually gained popularity and a growing research area of learning analytics has emerged to provide a more objective picture of student learning (Baker & Siemens, 2014).

Both theory-driven and data-driven approaches, however, have limitations: the theory-driven approaches are criticized for being subjective and lacking accuracy in self-assessment by subjects (Siemens, 2013); whereas the data-driven approaches are often fragmented from educational theories and rely purely on empiricism, which limit the insights they can offer for directing pedagogical reforms, guiding learning design, and enhancing student learning experiences and outcomes (Buckingham Shum & Crick, 2012). Theoretically speaking, it is useful to investigate in what ways the associations between students' learning processes and academic outcomes revealed by the two approaches are convergent or divergent. To address this issue, the study first investigated the assessments of students' academic performance by using their online learning patterns adopting a theory-driven or a data-driven approach separately. It then examined the extent to which the online learning patterns found in the two approaches are consistent with each other.

From a methodological point of view, the study combined methods used in theory-driven and data-driven approaches. Such combination not only provides complementary information of what students reported and what they actually did for online learning, but also enables the results obtained from each approach to be triangulated. It has the strengths of offering information regarding the intents of students' learning and objective evidence of their learning behaviors to address the limitations of adopting either a theory-driven or a data-centric approach (Reimann, Markauskaite, & Bannert, 2014).

In the context of a theory-driven approach, we applied Student Approaches to Learning (SAL) framework (Trigwell & Prosser, 2020) to investigate students' approaches to using online learning technologies and their perceptions of the online learning environment in the blended course through a self-reported Likert-scale questionnaire to demonstrate their self-reported online learning patterns. For the data-driven approach, we used the digital-trace data of the sequences of the online learning events produced by the learning analytic functions in the learning management system (LMS) to show students' online learning patterns by observation (Jovanović, Gašević, Pardo, Dawson, & Mirriahi, 2017). The following part will review the relevant SAL research and learning analytics research.

2. Literature review

2.1. Related SAL research

SAL is a well-established research framework into student learning in higher education (Trigwell & Prosser, 2020). It focuses on identifying various factors in the learning processes that are able to explain differences in academic performance (Biggs, Kember, & Leung, 2001). SAL research mostly uses self-reported questionnaires to examine variations of the ways how students go about their learning (i.e., approaches to learning) and how they perceive the learning environment and teaching context (i.e., perceptions of learning and teaching) (Ramsden, 2003). The approaches and perceptions in SAL research are considered being responses to different learning and teaching contexts rather than a personality trait (Diseth, 2003). Hence, the same individual may report adopting different approaches and having different perceptions from one environment to another (Biggs & Tang, 2011). In addition, individuals in the same learning context can report contrasting experiences, despite studying the same learning tasks and experiencing the same teaching team.

A key insight from SAL research is that consistent and logical relations have been found between students' approaches, perceptions, and their academic performance (Asikainen & Gijbels, 2017; Entwistle, 2009). Students adopting surface approaches to learning, which are characterized by mechanistic procedures, producing formulaic responses, and being not engaged with the conceptions in learning, tend to perceive poor teaching quality, high workload, and irrelevant assessments. In contrast, students who adopt deep approaches, which involve using meaningful strategies to investigate the subject matter and pursue in-depth understanding of the key concepts and theories, are likely to perceive teaching as fostering independence and clear goal-focused, workload and assessment tasks as appropriate (Lizzio, Wilson, & Simons, 2002; Wilson & Fowler, 2005).

In blended course designs, deep approaches to learning and to using online learning technologies are also found to be logically related to students' appreciation of the online design and valuing of online contributions, and relatively higher course marks; whereas surface approaches to learning and to using online learning technologies are significantly associated with perceptions of unbalanced online learning workload, and a disconnection between face-to-face and online parts of the course, and relatively lower academic learning outcomes (Ellis & Bliuc, 2016; Ellis & Bliuc, 2019; Han & Ellis, 2020a). In this study, a self-reported questionnaire from SAL research is used to examine students' online learning patterns, including their approaches to using online learning technologies and perceptions of the online learning environment in the blended course.

2.2. Related learning analytics research

Learning analytics research has been established at the intersection of learning sciences, computer science, psychology, and education (Gašević, Dawson, & Siemens, 2015). It focuses on how large-scale data derived from technologies can be used to increase the understanding and improvement of the quality of student learning experiences and outcomes (Siemens & Gašević, 2012). The large volume of digital-trace data records what students and teachers do when they interact with a variety of learning technologies. When combining with various students' demographic information, the big learning analytic data are processed by advanced data mining techniques and sophisticated algorithms so that they can be used to: (1) address challenging problems in education, such as identifying at-risk students to minimise course attrition (Krumm, Waddington, Teasley, & Lonn, 2014), increasing program retention (Dawson, Jovanović, Gašević, & Pardo, 2017), and monitoring students' affect (Ocumpaugh, Baker, Gowda, Heffernan, & Heffernan, 2014); (2) provide empirical evidence to support decision-making, like in the area of academic success prediction (Romero, López, Luna, & Ventura, 2013), education policy reforms (Ferguson et al., 2016), and career advice (Bettinger & Baker, 2013); and (3) improve learning and teaching quality, including assisting learning design (Tempelaar, Rienties, & Giesbers, 2015), identifying learning strategies (Chen, Resendes, Chai, & Hong, 2017), facilitating online discussions (Joksimović, Gašević, Kovanović, Riecke, & Hatala, 2015) and collaboration (Kaendler, Wiedmann, Rummel, & Spada, 2015), and providing personalised learning feedback (Pardo, Jovanović, Dawson, Gašević, & Mirriahi, 2019).

2.3. Combining theory-driven and data-driven approaches

In considering the drawbacks of relying solely on theory-driven or data-driven approaches, some researchers have proposed to combine theory-driven and data-driven approaches in explaining students' learning outcomes (Lockyer, Heathcote, & Dawson, 2013). A combination of theory-driven and data-driven approaches may not only increase the power of detection of learning behaviors and prediction of learning outcomes, but may also be used as a way of triangulation to see if the evidence from different approaches can achieve consistency (Rodríguez-Triana, Martínez-Monés, Asensio-Pérez, & Dimitriadis, 2015).

Adopting a combined approach, studies have used different sources of data to predict and assess students' learning outcomes. Some studies reported that student learning processes collected from different data sources contributed uniquely to the learning outcome and increased the predictive power (Han & Ellis, 2017a). For instance, Pardo, Han, and Ellis (2017) reported that adding the frequency of interactions with the online learning resources by observation explained an additional 25% of variance in students' course marks than using students' reported use of self-regulated learning strategies alone. Ellis, Han, and Pardo (2017) also found a similar result that inclusion of students' engagement with various online learning activities in the regression model increased as high as 25% of variance explained for students' course marks than merely using students' reported approaches to study.

However, research findings are not always consistent with regard to whether the self-reported and observational data of student learning complementarily explain students' academic performance or overlap. Some studies found either an indirect contribution (Pardo, Han, & Ellis, 2016) or non-significant contribution of self-reported data (Tempelaar et al., 2015) to learning outcomes after adding observational data. Using a path analysis, Pardo et al. (2016) showed that students' reported positive self-regulated learning strategies only indirectly predicted academic performance via online activity participation recorded by digital traces. Tempelaar et al. (2015) conducted a large-scale study, which included 151 online learning modules involving 11,256 students. Their regression analyses showed that students' reported satisfaction with the online learning modules became a non-significant predictor after entering the observational data of time spent on online learning, which explained 11% of variance of their module retention.

Noting these inconsistencies, researchers have attempted to triangulate evidence offered by theory-driven and data-driven approaches. Han and Ellis (2017b) reported logical associations between students' self-reported perceptions of the course learning environment and the recorded counts of students' use of online learning tools in the LMS. Positive perceptions were found to be related to higher counts of tool use; as well as positive associations between negative perceptions and lower counts of tool use, suggesting a consistency of the learning experiences obtained from the self-reported and observational data.

Using more complex observational data rather than frequency counts as in Han and Ellis (2017b), and Han, Pardo, and Ellis (2020) identified a self-reported "understanding" and a "reproducing" learning orientation by students' approaches to learning in face-to-face and online contexts and perceptions of the blended learning environment. They also identified four different online learning orientations by using the observational sequences of the online study states. The results showed that students who reported an "understanding" learning orientation were involved in more online study states with high volume of formative learning, indicating a level of consistency between learning orientations by self-reported and observational data. One limitation of this study is the mismatch between the self-reported data and the observational data: while the self-reported questionnaire measured the learning in both face-to-face and online parts in the course; the observational data only recorded the online learning. Such mismatch may affect the results and needs to be addressed in future research.

2.4. Research purposes and research questions

Combining theory-driven and data-driven approaches, the current study had three research purposes. The first two research purposes concerned with separate examinations of the self-reported and observational online learning patterns, and the relations between these patterns and the academic performance. While the self-reported online learning patterns were measured by students' reporting on approaches to online learning technologies and perceptions of the online learning environment; the observational online learning patterns were assessed by digital traces of sequences of the online learning events. The third research purpose was to investigate the level of consistency between the self-reported and observational online learning patterns. By combining theory-driven and data-driven approaches, the study aimed to improve the current practice of predicting learning outcomes in order to contribute to the development of precision education.

To be more specific, the study addressed three research questions:

- What are students' self-reported online learning patterns and their relations to the academic performance?
- What are students' observational online learning patterns and their relations to the academic performance?
- To what extent are the online and observational learning patterns consistent with each other?

3. Method

3.1. The participants and the course design

The research was conducted in an Australian metropolitan university with 314 freshmen studying a compulsory engineering course. The course lasted 13 weeks, and was designed with face-to-face and online components. The face-to-face component included weekly lectures (2 hours), weekly tutorials (2 hours), and weekly laboratory sessions (3 hours). The focus of the lectures was in-depth explanations of the key concepts, which were further expanded and discussed in tutorials. The tutorials also included activities of how to apply theoretical principles to tackling practical issues through demonstration of problem-solving tasks. In laboratory sessions, students were given opportunities to gain hand-on skills through projects, such building an electronic circuit, or configuring a computer system. The online component, which was held in a bespoke LMS, required students to engage with the online learning activities both before and after classes through three types of online learning activities, namely: (1) view course contents; (2) doing theory-related exercises; and (3) performing problem-solving exercises. Before classes, students were asked to familiarize themselves with the lectures, tutorials, and laboratory contents, such as key concepts and laboratory procedures through reading and/or watching videos (i.e., view course contents). After classes, there were quizzes to test students' understanding of terminologies, such as Moore's law, System Verilog, and Flynn's taxonomy (i.e., doing theory-related exercises). There were also mini case studies for students to solve practical problem sequences by applying theories, such as digital circuit design improvement, control circuit reset, and pipelined processor implementation (i.e., performing problem-solving exercises).

3.2. Data and the instruments

3.2.1. Self-reported data collected by a questionnaire

The self-report data was collected using a 5-point Likert-scale questionnaire, which consisted of five scales: two scales assessed approaches to using online learning technologies; and the other three scales evaluated perceptions of the online learning environment in the blended course, including perceptions of the integrated learning environment, perceptions of online contributions, and perceptions of online workload. These scales were developed adopting SAL framework (Prosser & Trigwell, 1999) and used and validated in previous SAL studies (Ellis & Bliuc, 2016; Han & Ellis, 2020b). The description and reliability of each scale accompanied by a sample item are provided in Table 1.

Table 1. The details of the questionnaire

Scale	Description	α	Sample
Deep approach to using online learning technologies (6 items)	Using online learning technologies in a meaningful way, such as assisting forming key inquiry questions in learning, deepening concepts in the course, and developing essential skills	.75	*I find I use the learning technologies in this course to further my research into a topic.*
Surface approach to using online learning technologies (5 items)	Using online learning technologies in formulaic and mechanistic ways, such as fulfilling course requirements and downloading documents	.75	*I only use the online learning technologies in this course to fulfil course requirements.*
Perceptions of the integrated learning environment (9 items)	Perceptions of levels of integration of the online learning in the course	.89	*The ideas we reviewed online helped with the assessment of the course.*
Perceptions of online contributions (6 items)	Perceptions of the value of online contributions by other students in the course	.87	*The online contributions from other students helped develop my understanding of particular topics.*
Perceptions of online workload (6 items)	Perceptions of the online workload in relation to the whole course	.77	*The workload for the online activities was too heavy. (negatively worded item)*

3.2.2. Observational digital-trace data collected by the learning analytic functions in the LMS

The observational digital-trace data was collected by the learning analytic functions in LMS, which recorded sequences of timestamped learning events involving three learning activities: (1) viewing course contents of printed and video materials; (2) doing theory-related exercises; and (3) performing problem-solving exercises.

3.2.3. Data of academic performance

The academic performance data was students' course marks, which were comprised by scores for class preparation (25%), the laboratory project (25%), and the close-book examination (50%). The examination assessed students' understanding of the key theories through 20 multiple-choice questions (1.5 marks each) and their abilities to apply theories in solving practical questions through four open questions (5 marks each). The range of the course marks were from 21.25 to 98.75 (M = 67.40, SD = 14.62).

3.3. Ethical consideration and data collection

We obtained the formal approval and strictly followed the ethical requirements stipulated by the institution's ethics committee. The volunteer students signed written consent forms and agreed to answer the questionnaire, permitted the extraction of the digital-trace data of their online learning, and allowed access to their course marks. The different types of the data were matched and then anonymized so that students' identification was revealed.

3.4. Data analysis methods

To answer the first research question, a hierarchical cluster analysis using the mean scores of the five scales of students' approaches to, and perceptions of, online learning was conducted to identify self-reported online learning patterns. Based on the cluster membership, one-way ANOVAs were performed to examine the relations between self-reported online learning patterns and academic performance. To facilitate interpretation, the mean scores of the scales were transformed into z-scores (M = 0, SD = 1) in the analyses. To provide an answer to the second research question, we performed an agglomerative sequence clustering analyses using the timestamped sequences of the online learning events to investigate the observational evidence of the online learning patterns. Then we applied a one-way ANOVA to see the extent to which students' academic performance differed by the observational online learning patterns. For the third research question, we run a cross-tabulation analysis using the self-reported and observational clusters generated from the above analyses.

4. Results

4.1. Self-reported online learning patterns and academic performance

A hierarchical cluster analysis using the five scales produced two clusters of students: cluster 1 had 95 students and cluster 2 had 219 students. As shown by one-way ANOVAs, the two clusters of students differed significantly on all the five scales: deep approach to using online learning technologies: $F(1, 312) = 88.43$, $p < .01$, $\eta^2 = .22$; surface approach to using online learning technologies: $F(1, 312) = 163.25$, $p < .01$, $\eta^2=.34$), perceptions of the integrated learning environment: $F(1, 312) = 119.53$, $p < .01$, $\eta^2 = .28$; perceptions of online contributions: $F(1, 312) = 8.59$ $p < .01$, $\eta^2 = .03$; and perceptions of online workload: $F(1, 312) = 115.60$, $p < .01$, $\eta^2 = .27$). The two clusters of students also differed on the course marks: ($F(1, 312) = 16.46$, $p < .01$, $\eta^2 = .05$); and the scores of each assessment task, class preparation: ($F(1, 312) = 4.11$, $p < .05$, $\eta^2 = .01$); laboratory project: ($F(1, 312) = 4.15$, $p < .05$, $\eta^2 = .01$); close-book examination: ($F(1, 312) = 14.92$, $p < .01$, $\eta^2 = .05$).

Table 2. Self-reported online learning patterns and academic performance

Variables	Deep (N = 95)		Surface (N = 219)		F	p	η^2
	M	SD	M	SD			
Deep approaches to online learning technologies	0.72	0.73	-0.31	0.96	88.43	.00	.22
Surface approaches to online learning technologies	-0.88	0.64	0.40	0.88	163.25	.00	.34
Perceptions of the integrated learning environment	0.80	0.62	-0.35	0.94	119.53	.00	.28
Perceptions of online contributions	0.24	0.95	-0.12	1.02	8.59	.00	.03
Perceptions of online workload	0.80	0.88	-0.34	0.86	115.60	.00	.27
Course marks	72.37	14.15	67.41	14.62	16.46	.00	.05
Class preparation	21.58	2.83	20.80	3.31	4.11	.04	.01
Laboratory project	21.48	6.17	19.84	6.71	4.15	.04	.01
Close-book examination	29.22	11.22	23.90	11.21	14.92	.00	.05

The M values in Table 2 suggested that cluster 1 students self-reported a higher score on deep approaches to using online learning technologies; positive perceptions of the integrated learning environment, of online contributions, and of online workload. These features of approaches and perceptions suggested that cluster 1 students had a "deep" learning pattern. In contrast, cluster 2 students reported higher scores on surface approaches to using online learning technologies, and had negative ratings on perceptions of the integrated learning environment, of online contributions, and of online workload. The learning of the cluster 2 students had characteristics of “surface” pattern of learning. The scores of assessment tasks achieved by students with the “deep” learning pattern were significantly higher than those with the “surface” learning pattern.

4.2. Observational online learning patterns and academic performance

The agglomerative hierarchical sequence clustering using the timestamped online learning events involving the three types of learning activities produced four observational online learning patterns, which are visualized in Figure 1.

- pattern 1 (N = 161): had most learning events (M learning events = 62), amongst which viewing course contents occupied highest proportion, followed by problem-solving exercises, and theory-related exercises accounted for the lowest proportion.
- pattern 2 (N = 64): had second most learning events (M learning events = 27), of which there were relatively balanced learning events of viewing course contents and doing theory-related exercises, with performing problem-solving exercises being lowest.
- pattern 3 (N = 27): had least learning events (M learning events = 13), of which there was predominantly doing theory-related exercises, with very low proportion of performing problem-solving exercises.
- pattern 4 (N = 62): had the second least learning events (M learning events = 18), of which there were high proportions of doing theory-related exercises, followed by viewing course contents, and performing problem-solving exercises had the lowest proportion.

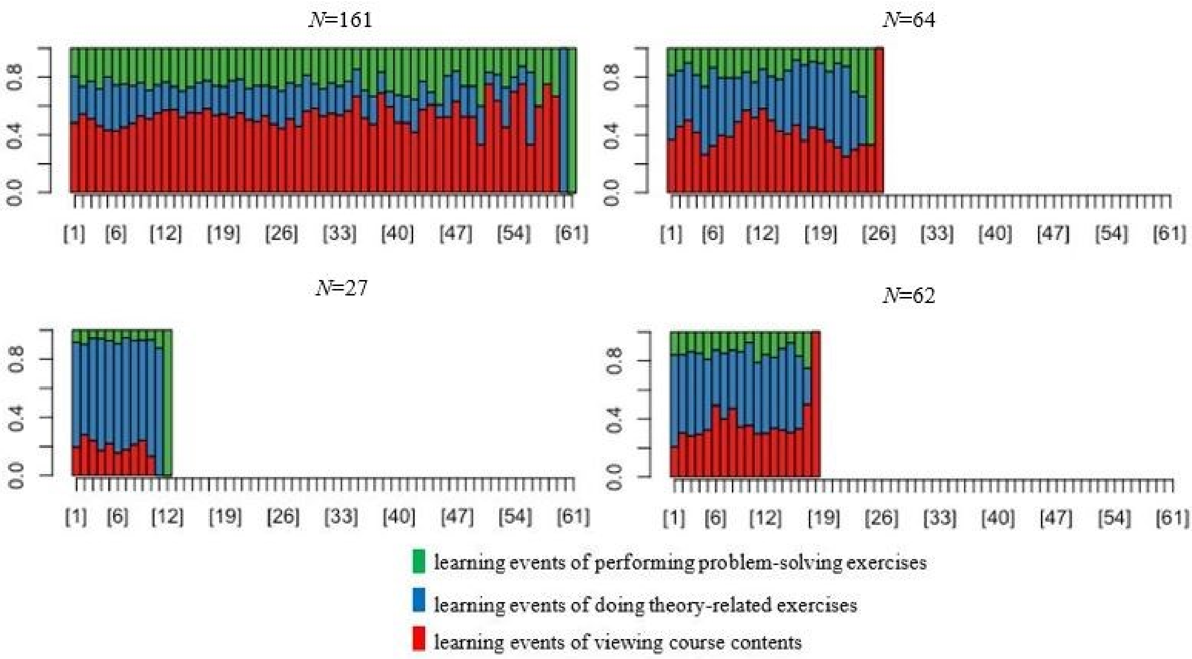


Figure 1. The four observational online learning patterns

The one-way ANOVAs showed that students’ course marks (F(3, 310) = 34.24, p < .01, η^2 = .25) and scores on each assessment task (class preparation: (F(3, 310) = 29.50, p < .01, η^2 = .22); laboratory project: (F(3, 310) = 12.96, p < .05, η^2 = .11); and close-book examination: (F(3, 310) = 15.87, p < .01, η^2 = .13) significantly differed by patterns. The post-hoc analyses in Table 3 were summarized (course marks: pattern1 > pattern2 > pattern3 = pattern4; class preparation: pattern1 > pattern2 > pattern3 > pattern4; laboratory project and close-book examination: pattern1 = pattern2 > pattern3 = pattern4).

Table 3. Post-hoc analyses of students' academic performance by patterns

	Course marks			*p* values for pairwise comparisons		
	N	*M*	*SD*	pattern 2	pattern 3	pattern 4
Pattern 1	161	73.13	12.99			
Pattern 2	64	68.39	13.22	.01		
Pattern 3	27	52.95	15.14	.00	.00	
Pattern 4	62	57.85	10.14	.01	.00	.10
	Class preparation			*p* values for pairwise comparisons		
	N	*M*	*SD*	pattern 2	pattern 3	pattern 4
Pattern 1	161	22.22	2.36			
Pattern 2	64	21.08	2.90	.01		
Pattern 3	27	17.61	3.78	.00	.00	
Pattern 4	62	19.41	3.34	.00	.00	.01
	Laboratory project			*p* values for pairwise comparisons		
	N	*M*	*SD*	pattern 2	pattern 3	pattern 4
Pattern 1	161	21.74	5.13			
Pattern 2	64	21.64	4.46	.92		
Pattern 3	27	17.41	7.26	.00	.00	
Pattern 4	62	16.62	9.25	.01	.00	.59
	Close-book examination			*p* values for pairwise comparisons		
	N	*M*	*SD*	pattern 2	pattern 3	pattern 4
Pattern 1	161	28.81	11.55			
Pattern 2	64	26.09	10.69	.09		
Pattern 3	27	17.78	11.02	.00	.00	
Pattern 4	62	19.72	8.04	.01	.00	.43

4.3. Association between self-reported and observational online learning patterns

The results of the cross-tabulation analysis revealed a significant association between the self-reported and observational learning patterns ($\chi^2(3) = 7.95$, $p < .05$). Table 4 shows that amongst the students categorized in observational online learning pattern 1, the proportion of students who reported a "deep" online learning pattern (61.1%) was significantly higher than that of students who reported a 'surface' online learning pattern (47.0%). Of the students in observational online learning pattern 3, the reversed pattern was observed: the proportion of students who reported a "surface" online learning pattern (11.0%) was significantly higher than that of those reporting a "deep" pattern (3.2%).

Table 4. Association between self-reported and observational online learning patterns

Patterns	Count & % within self-reported patterns	Deep	Surface	Total
Pattern 1	Count	58	103	161
	%	61.1% $_a$	47.0% $_b$	51.3%
Pattern 2	Count	18	46	64
	%	18.9% $_a$	21.0% $_a$	20.4%
Pattern 3	Count	3	24	27
	%	3.2% $_a$	11.0% $_b$	10.8%
Pattern 4	Count	16	46	62
	%	16.8% $_a$	21.0% $_a$	19.27%
Total	Count	95	219	314
	%	100.0%	100.0%	100.0%

Note. Different subscript letters denote the proportions differ significantly at $p < .05$.

5. Discussion and practical implications

An important aim of precision education is to provide tailored instructional interventions to address students' problematic learning behaviors in order to enhance their learning achievement (Cook et al., 2018; Tsai et al., 2020). Before effective intervention plans can be carried out, it is essential to know what students' problematic learning behaviors might entail and how they are related to learning outcomes. Understanding the structure of problematic student behaviours and why they do the things they do is one of the two foci of the current study.

Our results show that the different learning patterns identified by the self-reported approaches to, and perceptions of, online learning *and* the differences in the sequences of digital traces of the online learning events recorded in the LMS both are consistent with variations of the academic achievement as indicated by the final marks in their course.

Before discussing the results, it is worthwhile noting the limitations of the study. While the patterns of the observational online learning behaviors detected by the agglomerative sequence clustering considered the proportions of types of learning events as well as the total number of learning events, it did not provide detailed account of frequencies of each type of learning activities. Future studies could apply other statistical methods to make fine-grained analyses of the frequencies of different learning activities. Furthermore, due to ethical issues, it was not practical to obtain item-by-item score of students' close-book examination, hence, the internal reliability of the examination could not be calculated. Notwithstanding these limitations, the outcomes of this study offer some interesting insights into a possible way of improving prediction of students' learning outcomes from a precision education perspective.

In this study, the theory-driven approach found that students who self-reported deep approaches to using online learning technologies as well as positive perceptions of the online learning environment also obtained relatively higher achievement in the course; whereas those who reported using surface approaches and having negative perceptions were more likely to receive lower course marks. The two contrasting patterns of students' online learning were similar to the two contrasting patterns found in the research conducted in the traditional classroom learning context, which also found logical associations between deep approaches to learning, positive perceptions of teaching quality, and quality learning outcomes on the one hand; and surface approaches, perceptions of inappropriate workload and assessment, and poorer learning outcomes on the other (Lizzio et al., 2002; Wilson & Fowler, 2005). The results corroborated with the findings of the studies in blended course designs, which also distinguished between students' learning with contrasting approaches in both face-to-face and online parts, perceptions of integration between face-to-face and online learning, and the academic achievement (Ellis & Bliuc, 2019; Han & Ellis, 2020a). However, different to previous studies in blended contexts, the current study only focused on the approaches and perceptions of the online part in order to compare the observational data, which also concerned with the engagement with the online learning activities only. The comparison between the self-reported and observational data in our study improved the research design by addressing the limitation of a mismatch in Han's et al. (2020) research, in which the self-reported data focused on the whole course experience, whereas the observational data was only about online part of learning.

The self-reported findings show that as high as 70% of the students in the course did not approach learning in a meaningful and considered way. The questionnaire not only identified the students whose learning requires extra support, but also provided evidence to teachers to better understand which specific aspects of the learning experience should be improved so that the intervention strategies are more likely to produce a benefit. Specifically, the less desirable aspects of the learning experience in this study include adopting surface approaches to using online learning technologies, perceiving a fragmentation of the online learning experience in relation to the course, not valuing the online contributions of their peers, as well as considering the online workload to be high. Teachers can help improve these elements early in the course so that students may have more desirable learning experiences later on in the course. The teaching team can invite those whose learning was oriented towards a deep understanding of the subject matter to share their strategies and ideas. There are many useful topics for the students to share: such as how and why they used online learning technologies in a meaningful way to facilitate their learning, how these approaches helped them achieve the learning objectives in the course, what they decided to contribute online and what they found more useful to share in class, how they felt about and learnt from others' online contributions, how they managed their online workload and what proportion of time they allocated to their online learning activities in the overall course experience. In addition, the teachers can help all their students by providing more explicit explanations at the beginning of the course as to how online resources, quizzes, and activities are linked with learning outcomes to reduce the likelihood that students experience the online activities being unrelated to the learning objectives.

With the goal of discovering evidence to improve the students' experience of learning, it is equally valuable to employ a data-driven approach to find out the patterns of what, and how much, the students interacted with their online learning activities through investigating the digital traces left in the LMS. The results demonstrated that students who were involved in most learning events with higher proportions of tackling practical problems (pattern 1) tended to achieve relatively more highly compared to students, who participated the least in the learning events (pattern 4). The latter group also had a higher proportion of exercises focusing on just testing the understanding of theories and obtained the lowest course marks. The possible reasons as to why students in the observational online learning pattern 1 outperformed the students in the other patterns could be their active participation online, as reflected in the quantity of the completed online learning events. It could also be that the problem-solving exercises they favoured provided them opportunities to link theoretical concepts with practical problems, enabling a deeper understanding of the key theories and strengthening their understanding between the subject of learning and its applications in context. In addition, the practice of a higher proportion of the problem-solving type of online learning activities by the students adopting pattern 1 were congruent with the major assessment in the course, accounting for 50% of the course marks. The examination not only tested students' theoretical understanding but also their ability to solve practical problems by applying the theories. In contrast, the students in pattern 4 not only had insufficient practice, but the theory-related exercises they prioritized to work with lacked sophistication to allow them to extend theories into practice.

Difference in the depth of engagement with the online learning activities was also reflected by differences in the self-reported approaches and perceptions. The significant association between the self-reported and observational online learning patterns suggest that what students reported they did in the learning was consistent and coherent with what the observed data suggested they did. This outcome provides a type of triangulated evidence for both approaches (Knight, Buckingham Shum, & Littleton, 2014). The positive association found in the current study are also in line with a trend of associations identified in related research (Han & Ellis, 2017b; Han et al., 2020). This study has added to the quality of the research design by matching the self-reported and observational data so that both emphasised how the online part of learning experience in the blended course design. Interestingly, the results of our study are only partially consistent with the study by Gašević, Jovanović, Pardo, and Dawson (2017), which found significant link between self-reported and observational learning approaches only for the deep aspect but not for the surface aspect. Clearly more studies are needed before more conclusive evidence can be drawn.

The results from the data-drive approach suggest to us that teachers can monitor both the quantity and the quality of students' engagement with the online learning activities through using the learning analytics functions built in the LMS. The early detection of such information can hint to teachers the levels of engagement and the appropriateness of strategies the students approach their online learning in order to implement the class-level and/or personalized intervention strategies. For the whole class, the teaching staff can encourage online participation by selecting some online activities as part of discussions in the class to make these as essential activities for students to prepare beforehand. Teachers can also explain the purpose of each type of the online learning activities and their relations to help students make wise decision as to how much time and effort to be allocated for each one. Personalized intervention strategies could be also arranged to support students with different problems in learning. For example, the dashboard of students' online participation together with the class average online participation rate could be sent to those at the bottom level of online engagement to reminder them of catching up. Personalized strategic plans for how to deal with the online activities in an appropriate way can also be tailored and delivered to those lacking good strategies. It is hoped that through those timely interventions, students can make adjustments and improve their learning behaviors, lowering the potential risk of dropping out or failing the course.

6. Conclusion and implications for precision education

Research in precision education in an era of big data often solely relies on the techniques of advanced statistical learning, deep learning, and sophisticated data mining to achieve the "volume" principle in the four V's of the big data analytics (i.e., volume, velocity, variety, and veracity; IBM, 2018). This often results in the disintegration between the quantitative numbers and the established educational theories. Hence, the figures and the models generated in datacentric approach in precision education may severely limit the power and value in the abilities of guiding practice in learning and teaching or translating the numbers into meaningful interpretations for stakeholders in policy-making processes (Chan, 2019). Therefore, within current international trends in educational research, it is timely to combine theory-driven and data-driven approaches to advance the applications of big data analytics in precision education, so that more attention can be paid to the 'variety' principle to include multiple data sources, data collection methods, and data processing techniques in a single study (Topps & Cullen, 2019).

Our study is an initial attempt to demonstrate how combining theory-driven and data-driven approaches can be used to improve research on predicting students' academic achievement, which is one of the major objectives in precision education. Our study predicted learning outcomes by using students' perceived online learning experience through self-reports and their actual online learning behaviors observed by the digital-trace data in LMS. The strengths of such combined approaches lie in multiple ways. First, it has an advantage of offering richer information in the way of predicting students' learning over using a single approach, with each approach supplementing the other. While the observational digital-trace data are able to provide objective evidence as to what students actually do in their learning (Fincham, Gašević, Jovanović, & Pardo, 2019), they do not, however, have capacity to reflect students' intents and perceptions behind the ways they learn as in the self-reported studies (Asikainen & Gijbels, 2017; Gerritsen-van Leeuwenkamp, Joosten-ten Brinke, & Kester, 2019). Second, combined approaches can serve as a triangulation to check the validity of the results derived from either a theory-driven or data-driven approach. The significant and logical association found between the learning patterns of the two types of data in our study demonstrate complementarity and some degree of consistency and coherence of the two approaches. Third, the multiple data analysis methods used in the combined approaches also strengthen the analytical power of the analyses. Hence, the combined approaches have potential to transfer into other similar investigations, which tackle the complex issues of contemporary student experiences of learning, involving interactions not only between students and other individuals (e.g., peers, teachers, tutors, and laboratory facilitators), but also between students and things (e.g., tools, resources, and learning spaces). All the merits of combining theory-driven and data-driven approaches point out its future applications to advance research in precision education. While the current study demonstrates how combined approaches are used to improve current practice of predicting students' learning outcomes using either a theory-driven or data-driven perspective, future studies may extend this methodology to fulfil other objectives of precision education, such as diagnosing learning problems, personalized learning interventions, and preventing learning failure.

Acknowledgement

The authors are pleased to acknowledge the financial support of the Australian Research Council through grant DP150104163.

References

Asikainen, H., & Gijbels, D. (2017). Do students develop towards more deep approaches to learning during studies? A Systematic review on the development of students' deep and surface approaches to learning in higher education. *Educational Psychology Review, 29*(2), 205-234.

Baker, R., & Siemens, G. (2014). Educational data mining and learning analytics. In R. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 253-272). New York, NY: Cambridge University Press.

Bettinger, E., & Baker, R. (2013). The Effects of student coaching: An Evaluation of a randomized experiment in student advising. *Educational Evaluation and Policy Analysis, 36*(1), 3-19.

Biggs, J., Kember, D., & Leung, D. (2001). The Revised two-factor study process questionnaire: R-SPQ-2F. *British Journal of Educational Psychology, 71*(1), 133-149.

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university: What the student does* (4th ed.). London, UK: McGraw-Hill.

Bliuc, A.-M., Goodyear, P., & Ellis, R. A. (2007). Research focus and methodological choices in studies into students' experiences of blended learning in higher education. *The Internet & Higher Education, 10*(4), 231-244.

Buckingham Shum, S., & Crick, R. (2012). Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 92-101). New York, NY: ACM.

Chan, D. (2009). So why ask me? Are self-report data really that bad? In C. Lance & R. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 309-336). New York, NY: Routledge.

Chen, B., Resendes, M., Chai, C., & Hong, H. (2017). Two tales of time: Uncovering the significance of sequential patterns among contribution types in knowledge-building discourse. *Interactive Learning Environments, 25*(2), 162-175.

Collins, F., & Varmus, H. (2015). A New initiative on precision medicine. *New England Journal of Medicine, 372*(9), 793-795.

Cook, C., Kilgus, S., & Burns, M. (2018). Advancing the science and practice of precision education to enhance student outcomes. *Journal of School Psychology, 66*, 4-10.

Dawson, S., Jovanović, J., Gašević, D., & Pardo, A. (2017). From prediction to impact: Evaluation of a learning analytics retention program. In X. Ochoa & I. Molenaar (Eds.), *Proceedings of the International Conference on Learning Analytics and Knowledge* (pp. 474-478). Vancouver, Canada: ACM.

Diseth, Å. (2003). Personality and approaches to learning as predictors of academic achievement. *European Journal of Personality, 17*(2), 143-155.

Ellis, R. A., & Bliuc, A.-M. (2016). An Exploration into first year university students' approaches to inquiry and online learning technologies in blended environments. *British Journal of Educational Technology, 47*(5), 970-980.

Ellis, R. A., & Bliuc, A.-M. (2019). Exploring new elements of the student approaches to learning framework: The Role of online Han, F., & Ellis, R. A. (2019). Identifying consistent patterns of quality learning discussions in blended learning. *The Internet & Higher Education, 40*, 12-19.

Ellis, R. A., & Goodyear, P. (2016). Models of learning space: Integrating research on space, place and learning in higher education. *Review of Education, 4*(2), 149-191.

Ellis, R. A., & Goodyear, P. (2019). *The Education ecology of universities: Integrating learning, strategy and the academy*. New York, NY: Routledge.

Ellis, R. A., Han, F., & Pardo, A. (2017). Improving insights from learning analytics – The Value of combining observed and self-report data on university student learning. *Educational Technology & Society, 20*(3), 158-169.

Entwistle, N. (2009). *Teaching for understanding at university: Deep approaches and distinctive ways of thinking.* Eastbourne, UK: Palgrave Macmillan.

Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittlemeier, J., Rienties, B., Ullmann, T., & Vuorikari, R. (2016). Research evidence on the use of learning analytics: Implications for education policy. Seville, Spain: Joint Research Centre. doi:10.2791/955210

Fincham, E., Gašević, D., Jovanović, J., & Pardo, A. (2019). From study tactics to learning strategies: An Analytical method for extracting interpretable representations. *IEEE Transactions on Learning Technologies, 12*(1), 59-72.

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64-75.

Gašević, D., Jovanović, J., Pardo, A., & Dawson, S. (2017). Detecting learning strategies with analytics: Links with self-reported measures and academic performance. *Journal of Learning Analytics, 4*(2), 113-128.

Gerritsen-van Leeuwenkamp, K., Joosten-ten Brinke, D., & Kester, L. (2019). Students' perceptions of assessment quality related to their learning approaches and learning outcomes. *Studies in Educational Evaluation, 63*, 72-82.

Hadwin, A., & Oshige, M. (2011). Self-regulation, coregulation, and socially shared regulation: Exploring perspectives of social in self-regulated learning theory. *Teachers College Record, 113*(2), 240-264.

Han, F., & Ellis, R. A. (2017a). Explaining learning achievement in student experience of blended learning: What can a sociomaterial perspective contribute? In H. Partridge, K. Davis, & J. Thomas. (Eds.), *Proceedings of the 34th International Conference on Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education* (pp. 89-94). Toowoomba, Australia: University of Southern Queensland.

Han, F., & Ellis, R. A. (2017b). Variations in coherence and engagement in students' experience of blended learning. In H. Partridge, K. Davis, & J. Thomas. (Eds.), *Proceedings of the 34th International Conference on Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education* (pp. 268-275). Toowoomba, Australia: University of Southern Queensland.

Han, F., & Ellis, R. A. (2020a). Personalised learning networks in the university blended learning context. *Comunicar, 62*(1), 19-30.

Han, F., & Ellis, R. A. (2020b). Initial development and validation of the Perceptions of the Blended Learning Environment Questionnaire. *Journal of Psychoeducational Assessment, 38*(2) 168-181.

Han, F., Pardo, A., & Ellis, R. A. (2020). Students' self-report and observed learning orientations in blended university course design: How are they related to each other and to academic performance? *Journal of Computer Assisted Learning, 36*(6), 969-980.

Hart, S. (2016). Precision education initiative: Moving toward personalized education. *Mind, Brain, and Education, 10*(4), 209-211.

IBM. (2018). *The Four v's of big data. IBM Big Data and Analytics Hub.* Retrieved from https://www.ibmbigdatahub.com/infographic/four-vs-big-data

Joksimović, S., Gašević, D., Kovanović, V., Riecke, B., & Hatala, M. (2015). Social presence in online discussions as a process predictor of academic performance. *Journal of Computer Assisted Learning, 31*(6), 638-654.

Jovanović, J., Gašević, D., Pardo, A., Dawson, S., & Mirriahi, N. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *The Internet & Higher Education, 23*, 74-85.

Kaendler, C., Wiedmann, M., Rummel, N., & Spada, H. (2015). Teacher competencies for the implementation of collaborative learning in the classroom: A Framework and research review. *Educational Psychology Review, 27*(3), 505-536.

Knight, S., Buckingham Shum, S., & Littleton, K. (2014). Epistemology, assessment, pedagogy: Where blearning meets analytics in the middle space. *Journal of Learning Analytics, 1*(2), 23-47.

Krumm, A., Waddington, R., Teasley, S., & Lonn, S. (2014). A Learning management system-based early warning system for academic advising in undergraduate engineering. In J. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 103-119). New York, NY: Springer.

Laurillard, D. (2013). *Rethinking university teaching: A Conversational framework for the effective use of learning technologies.* London, UK: Routledge.

Lemons, C., King, S., Davidson, K., Puranik, C., Al Otaiba, S., & Fidler, D. (2018). Personalized reading intervention for children with Down syndrome. *Journal of School Psychology, 66*, 67-84.

Lizzio, A., Wilson, K., & Simons, R. (2002). University students' perceptions of the learning environment and academic outcomes: Implications for theory and practice. *Studies in Higher Education, 27*, 27-52.

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist, 57*(10), 1439-1459.

López-Pérez, M., López-Pérez, M., & Rodríguez-Ariza, L. (2011). Blended learning in higher education: Students' perceptions and their relation to outcomes. *Computers & Education, 56*(3), 818-826.

Lu, O., Huang, A., Huang, J., Lin, A., Ogata, H., & Yang, S.-J.-H. (2018). Applying learning analytics for the early prediction of students' academic performance in blended learning. *Educational Technology & Society, 21*(2), 220-232.

Ocumpaugh, J., Baker, R., Gowda, S., Heffernan, N., & Heffernan, C. (2014). Population validity for educational data mining models: A case study in affect detection. *British Journal of Educational Technology, 45*(3), 487-501.

Pardo, A., Han, F., & Ellis, R. A. (2016). Exploring the relation between self-regulation, online activities, and academic performance: A Case study. In *Proceedings of the 6th International Learning Analytics and Knowledge Conference* (pp. 422-429). Edinburgh, UK: ACM Press.

Pardo, A., Han, F., & Ellis, R. A. (2017). Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance. *IEEE Transactions on Learning Technologies, 10*(1), 82-92.

Pardo, A., Jovanović, J., Dawson, S., Gašević, D., & Mirriahi, N. (2019). Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology, 50*(1), 128-138.

Prosser, M., & Trigwell, K. (1999). *Understanding learning and teaching. The Experience in higher education.* London, UK: McGraw-Hill.

Ramsden, P. (2003). *Learning to teach in higher education* (2nd ed.). London, UK: Routledge.

Reimann, P., Markauskaite, L., & Bannert, M. (2014). e-Research and learning theory: What do sequence and process mining methods contribute? *British Journal of Educational Technology*, *45*(3), 528-540.

Rodríguez-Triana, M., Martínez-Monés, A., Asensio-Pérez, J., & Dimitriadis, Y. (2015). Scripting and monitoring meet each other: Aligning learning analytics and learning design to support teachers in orchestrating CSCL situations. *British Journal of Educational Technology, 46*(2), 330-343.

Rojas-López, A., & García-Peñalvo, F. (2019). Personalized education for a programming course in higher education. In M. Sein-Echaluce, Á. Fidalgo-Blanco, & F. García-Peñalvo (Eds.), *Innovative trends in flipped teaching and adaptive learning* (pp. 203-227). Hershey, PA: IGI.

Romero, C., López, M.-I., Luna, J.-M., & Ventura, S. (2013). Predicting students' final performance from participation in on-line discussion forums. *Computers & Education, 68*, 458-472.

Siemens, G. (2013). Learning analytics: The Emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380-1400.

Siemens, G., & Gašević, D. (2012). Guest editorial: Learning and knowledge analytics. *Educational Technology & Society, 15*(3), 1-2.

Tempelaar, D., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior, 47*, 157-167.

The White House (2015). *Fact sheet: President Obama's precision medicine initiative*. Retrieved from http://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative

Topps, D., & Cullen, M. (2019). Scalable use of big data analytics in small group learning. *MedEdPublish*. doi:10.15694/mep.2019.000138.1

Trigwell, K. & Prosser, M. (2020). *Exploring university teaching and learning: Experience and context.* London, UK: Palgrave Macmillan.

Tsai, S., Chen, C., Shiao, Y., Ciou, J., & Wu, T. (2020). Precision education with statistical learning and deep learning: A Case study in Taiwan. *International Journal of Educational Technology in Higher Education, 17*, 1-13.

Williamson, B. (2019). Digital policy sociology: Software and science in data-intensive precision education. *Critical Studies in Education,* 1-17. doi:10.1080/17508487.2019.1691030

Wilson, K., & Fowler, J. (2005). Assessing the impact of learning environments on students' approaches to learning: Comparing conventional and action learning designs. *Assessment and Evaluation in Higher Education, 30*, 85-99.

Wilson, M., & Ismaili, P. (2019). Toward maximizing the student experience and value proposition through precision education. *Business Education Innovation Journal, 11*(2) 119-124.

Wu, J.-H., Tennyson, R., & Hsia, T.-L. (2010). A Study of student satisfaction in a blended e-learning system environment. *Computers & Education, 55*(1), 155-164.

Yang, S. J. H. (2019). Precision education: New challenges for AI in education [conference keynote]. In *Proceedings of the 27th International Conference on Computers in Education (ICCE)* (pp. XXVII-XXVIII). Kenting, Taiwan: Asia-Pacific Society for Computers in Education (APSCE).

Chen, X., Zou, D., Xie, H., & Cheng, G. (2021). Twenty Years of Personalized Language Learning: Topic Modeling and Knowledge Mapping. *Educational Technology & Society*, *24* (1), 205-222.

Twenty Years of Personalized Language Learning: Topic Modeling and Knowledge Mapping

Xieling Chen[1], Di Zou[2*], Haoran Xie[3], and Gary Cheng[1]
[1]Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong SAR // [2]Department of English Language Education, The Education University of Hong Kong, Hong Kong SAR // [3]Department of Computing and Decision Sciences, Lingnan University, Hong Kong SAR //
xielingchen0708@gmail.com // dizoudaisy@gmail.com // hrxie2@gmail.com // chengks@eduhk.hk
[*]Corresponding author

ABSTRACT: Personalized language learning (PLL), a popular approach to precision language education, plays an increasingly essential role in effective language education to meet diverse learner needs and expectations. Research on PLL has become an active sub-field of research on technology-enhanced language learning and artificial intelligence applications in education. Based on the PLL literature from the Web of Science and Scopus databases, this study identified trends and prominent research issues within the field from 2000 to 2019 using structural topic modeling and bibliometrics. Trend analysis of articles demonstrated increasing interest in PLL research. Journals such as *Educational Technology & Society* and *Computers & Education* had contributed much to PLL research. PLL associated closely with mobile learning, game-based learning, and online/web-based learning. Moreover, personalized feedback and recommendations were important issues in PLL. Additionally, there was an increasing interest in adopting learning analytics and artificial intelligence in PLL research. Results obtained could help practitioners and scholars better understand the trends and status of PLL research and become aware of the hot topics and future directions.

Keywords: Personalized language learning, Topic modeling, Knowledge mapping, Bibliometrics, Precision education

1. Introduction

Recent changes in curriculum design and pedagogical approaches emphasize the significance and effectiveness of personalized education in comparison to conventional cohort-based learning. Personalized education is one mode of precision education. They both consider individual differences to trigger the most effective intervention to meet the unique needs of individual learners and can identify at-risk students at early stages and provide timely intervention (Lu et al., 2018; Yang, 2019). Precision education involves the wide use of techniques of personalized learning, including learning analytics (LA) and adaptive learning software. It has been applied to various subjects and different education levels, with positive outcomes being reported. Connor et al. (2018) evaluated the efficacy of ISIMath, which tailored mathematics instruction for second-grade students. ISIMath significantly improved students' performance through individualized mathematics instruction based on assessment data. Chrysafiadi and Virvou (2013) presented ELaC, which provided adapted instructional materials based on the backgrounds, skills, and learning paces of individual learners. ELaC improved the adaptation efficiency of the instructional process and enhanced learning with personalized content and learning pace.

Personalized language learning (PLL) plays an important role in precision language education. It "heralds a new way of dealing with individual differences by effecting as precise a diagnosis as possible on each language learner, thus triggering specific interventions designed to target and respond to each person's specific language-learning problems" (Lian & Sangarun, 2017, p. 1). According to Lian and Sangarun (2017), personalization is the starting point to identify learner needs and provide precise solutions to satisfy their needs. Thus, personalization is "a subset of precision education", and precision education is "the ultimate objective" (p. 6). In other words, PLL is an important approach to precision language education.

Based on the definition of personalized learning by the US Department of Education (2017), this research defines PLL as an instruction that optimizes the pace, approaches, objectives, content, and activities of learning according to the interests and needs of individual language learners. Advances in analytic innovations and adaptive learning technologies have significantly facilitated the personalization of teaching and learning. Driven by the continuously growing requirement for the individualization of language learners' learning processes in a democratizing and globalizing world of exponential linguistic and cultural demands, PLL has become a prevailing focus of the educational technology industry, as well as a new challenge of the applications of artificial intelligence (AI), machine learning, and LA. Affordances of PLL have been highlighted. Wu et al. (2014) proposed a ubiquitous personalized English reading system based on RFID technology. The system recommended English articles with realistic scenarios to learners by analyzing their locations. Specifically, the system detected a learner's location and sent situation-related English articles for him/her to read and study. By considering the local context, the English content became more perceivable, thus supporting personalized and situational learning. Fang et al. (2018) proposed a content-driven method to recommend personalized grammar questions using a parse-key tree to detect grammatical structure and grammar question usage. The proposed approach effectively recommended grammar questions by considering both the conceptual and textual information of grammar questions.

A small number of review studies on personalized learning had been conducted. A representative one by Xie et al. (2019) discussed the status and tendencies of technology-assisted personalized/adaptive learning by reviewing 70 articles. Their study revealed that data sources such as students' profiles and learning logs were commonly used to support personalized/adaptive learning. Personalized/adaptive learning strategy had been integrated into many potential applications supported by smart devices and advances in AI, virtual reality, and wearable computing. Currently, few PLL reviews are available, with only one (Ismail et al., 2016) focusing on the classification, trends, and challenges of PLL systems. Their study suggested that PLL systems could be further improved by incorporating more complex adaptive learner models and contextualized learning tasks. Recognizing the significance of PLL research, a thorough analysis of the literature is needed to answer questions such as "what were the major issues in PLL" and "what is the future of PLL research." Such analysis can provide a state-of-the-art understanding of PLL research hotspots and useful implications for its future development.

Bibliometric analysis involves the application of mathematical/statistical techniques and quantitative measurements to evaluate academic literature (Chen et al., 2020a; Chen, et al., 2020d; Chen et al., 2018; Hao et al., 2018). Structural topic modeling (STM), a semi-automatic quantitative text mining approach with the basis of unsupervised machine learning, is receiving popularity among social science scholars "to discover topics from the data rather than assume them" (Roberts et al., 2014, p. 106). By combining STM with bibliometric analysis, this study analyzed PLL articles in terms of the trend of annual articles, top journals, countries/regions, institutions, essential research issues, and their evolutions to enable an in-depth understanding of the status and trends of PLL research. Findings obtained will enable scholars, educators, policymakers, and practitioners to better understand the latest PLL research and its developmental tendencies and to further facilitate its future development. Specifically, the following six major research questions were addressed:

(1) What was the trend of the annual number of PLL research articles?
(2) What were the top journals, countries/regions, and institutions ranked by Hirsch index (H-index)?
(3) What was the scientific co-authorship among major countries/regions and institutions?
(4) What were the major research foci?
(5) How did these research foci evolve?
(6) What were the research concerns of the major countries/regions and institutions?

2. Dataset and methods

The data collection and analysis flowchart (see Figure 1) includes data identification, data screening, and data analysis. The current study used bibliographic data collected from Web of Science (WoS) and Scopus databases using a search query as ((TS = ((personalization OR personalisation OR personalized OR personalised) AND (language) AND (learn*))) AND (Year of publication = 2000-2019) AND (Article type = journal articles)). In the query, "TS" (Topics) refers to the title, abstract, or keywords of a publication. The above search terms were decided with reference to previous work (e.g., "language" and "learning" in Zhang and Zou (2020) and "personalized learning" in Xie et al. (2019)) by considering both personalized learning and language learning. As indicated (Zhang et al., 2020), it was around 2004 that education systems worldwide were making efforts to personalize learning. To guarantee a full cover of PLL studies, we initially set the time span as recent two decades, i.e., from 2000 to 2019. We used journal articles because they usually undergo a meticulous peer-review process and are generally of high quality.

The following types of information were collected: titles, years of publication, authors and their institutions, and abstracts. With 650 initially retrieved articles, after excluding 159 duplicated articles, manual screening of the remaining 491 articles was conducted to ensure data relevance. The specific exclusion criteria for the screening are presented in Figure 1. When we decided whether a paper should be included, we started from the first criterion (i.e., relevant to language learning). If it was not, we excluded it directly without checking other criteria. In this first stage, a total of 314 articles were excluded, most of which were about the learning of programming languages. Moreover, many were about the use of machine learning methodology or natural language processing (NLP) that contain "learning" and "language" but were not about language learning by real learners. Subsequently, we read the article to check whether it was about personalized learning and included details concerning the PLL process. Some studies mentioned PLL as future research recommendations, while the main research per se was not about personalized learning. In this second stage, 40 articles were excluded. After confirming that the paper was related to PLL, we evaluated whether it was an original research article and excluded those that were reviews or survey papers. In this stage, 20 papers were excluded. Lastly, we evaluated whether the papers were on teacher education and excluded nine papers in this stage. The screening resulted in 108 relevant articles. The citation for each of them was provided by Google Scholar (see https://scholar.google.com/).

We then analyzed the 108 PLL articles regarding the trend of articles, top journals and contributors, scientific collaborations, and major research topics. Analysis methods included descriptive statistics, bibliometric indicators such as the H-index, social network analysis, STM, and the Mann-Kendall trend test. The STM was conducted using software R with title and abstract information as input data.

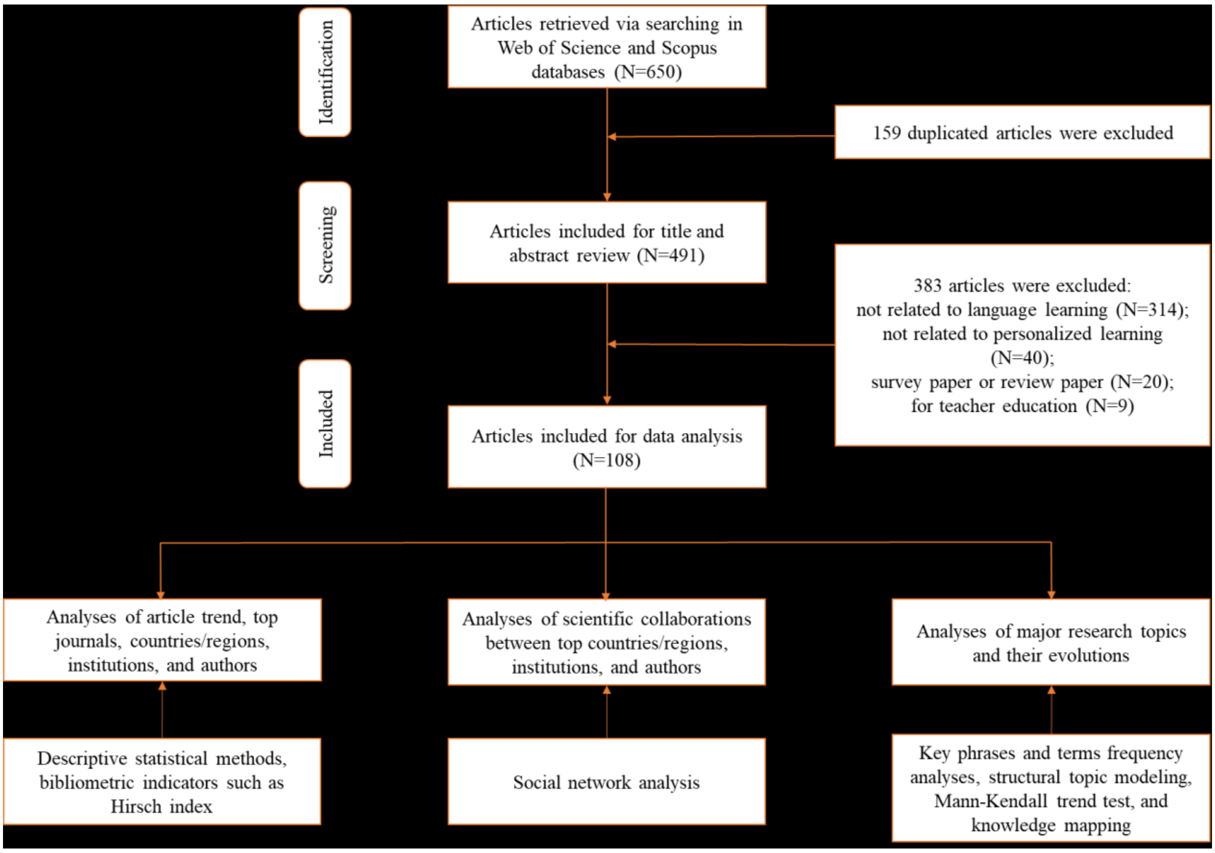


Figure 1. Flowchart of data collection and analyses

3. Results

3.1. Analysis of the trend of articles

Figure 2 shows the annual trend of PLL research. Overall, the annual number experienced an increasing trend from two in 2001 to 17 in 2019, demonstrating a constant increase in interest in PLL research, particularly since 2007. It is reasonable to anticipate that research enthusiasm in PLL will continue to increase in the future.

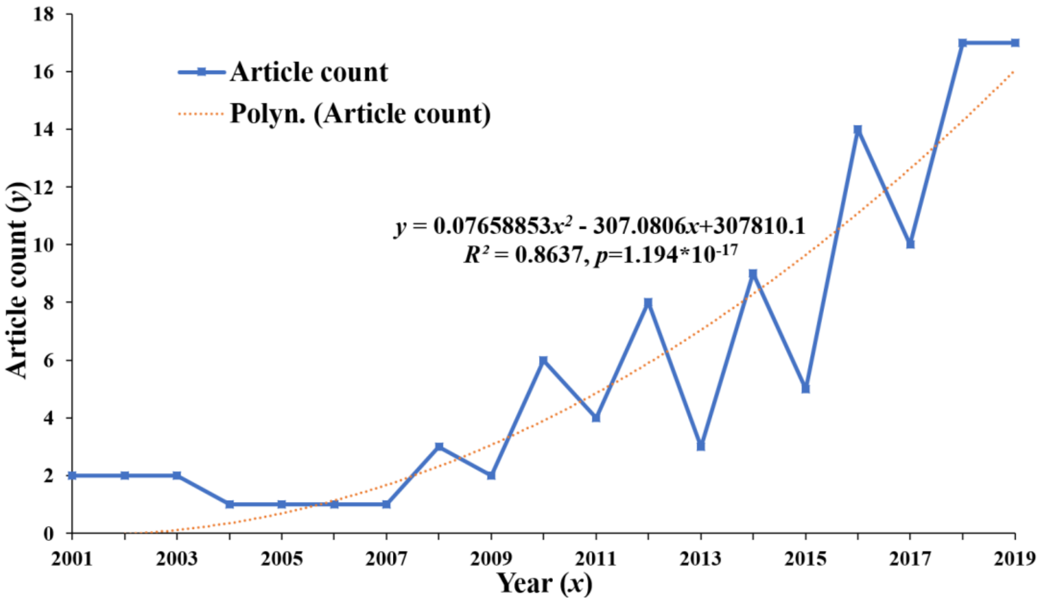


Figure 2. Year-by-year analysis of PLL studies

3.2. Top journals, countries/regions, and institutions

The 108 PLL articles were distributed in 77 journals. The top ones ranked by H-index (see Figure 3) accounted for 34.26% of the total articles. The top three were *Educational Technology & Society*, *Computers & Education*, and *Computer Assisted Language Learning*. The first two publish research about the application of technologies in education, and the last one specializes in applying technologies in language education. Among the listed journals, five are related to technology-enhanced language learning (i.e., *Computer Assisted Language Learning*, *Language Learning & Technology*, *Language Learning Journal*, *ReCALL*, and *System*). Meanwhile, over half of them are education-related, indicating a broad interest in PLL among education researchers, rather than limited to language researchers.

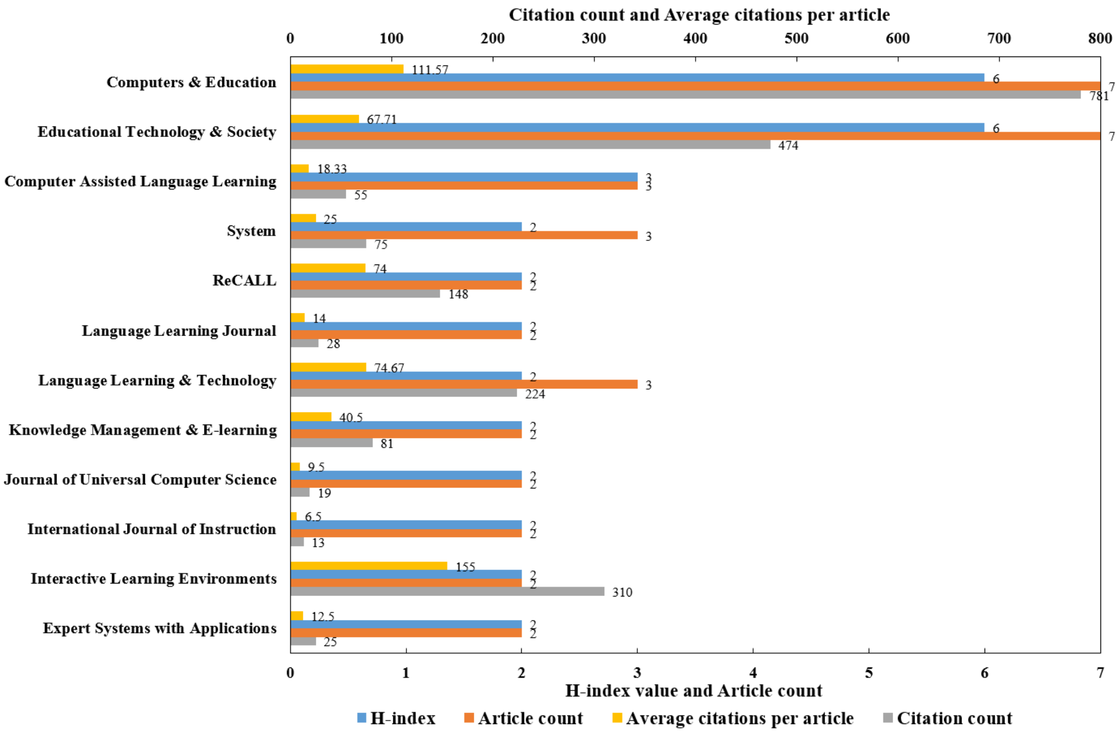


Figure 3. Top journals

There were 34 countries/regions and 147 institutions. Figure 4 presents the top 12 countries/regions ranked by H-index, indicating the important role of researchers in the Asia-Pacific region. Figure 5 presents the top institutions ranked by H-index. National Chengchi University was the most influential and prolific. Additionally, half of the top institutions are from Taiwan, indicating its dominance in PLL research.

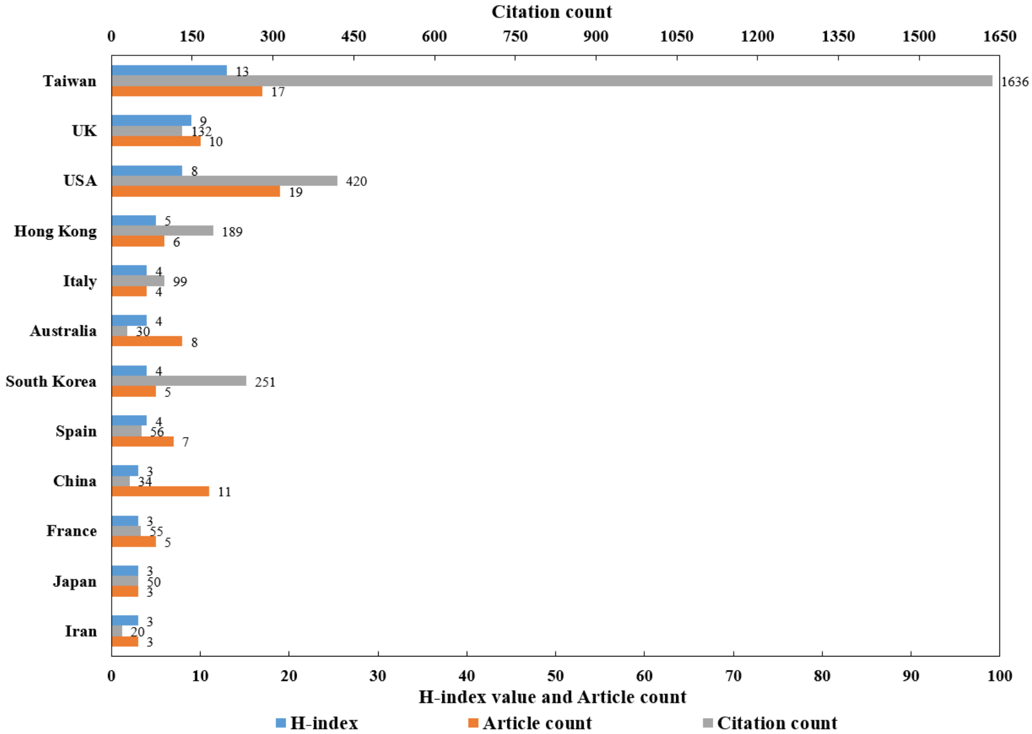


Figure 4. Top countries/regions

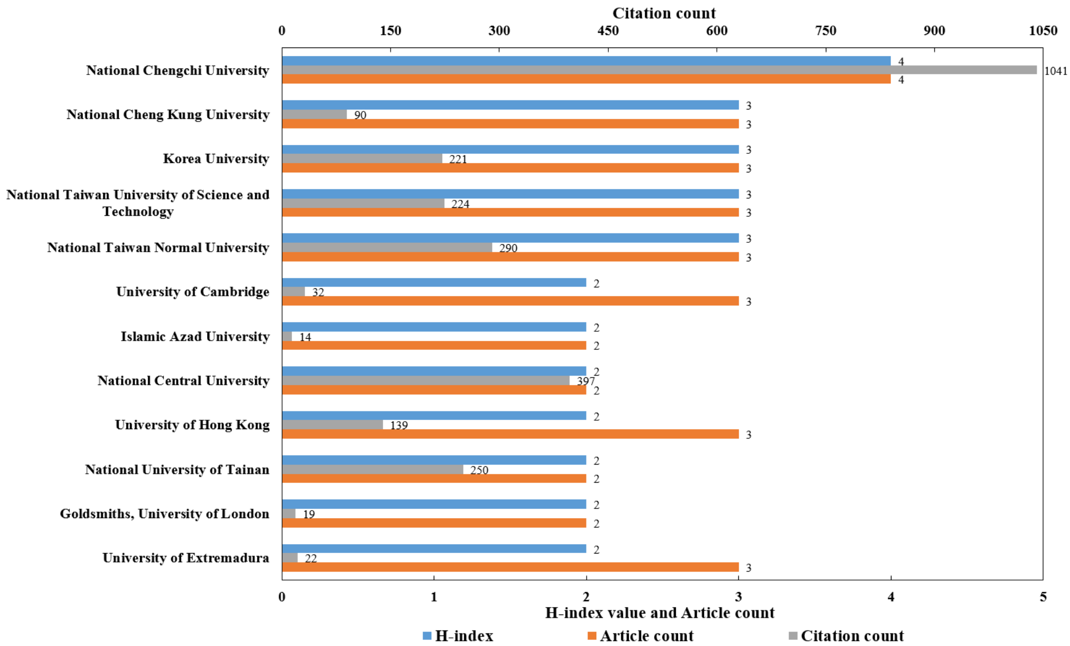


Figure 5. Top institutions

3.3. Analyses of scientific collaborations

The collaborations among the 34 countries/regions and the top 19 prolific institutions were visualized in Figures 6 and 7. Countries/regions and institutions were indicated using nodes with the size indicating the article count. Each node was colored based on its continental or national/regional information. Figure 6 shows that top collaborative partners included Belgium and the UK, Spain and the UK, as well as Hong Kong and China.

Collaborations among Asian and European regions were also close. From an institutional perspective, closest collaborative partners were also indicated in Figure 7, for example, East China Normal University and University of Hong Kong.

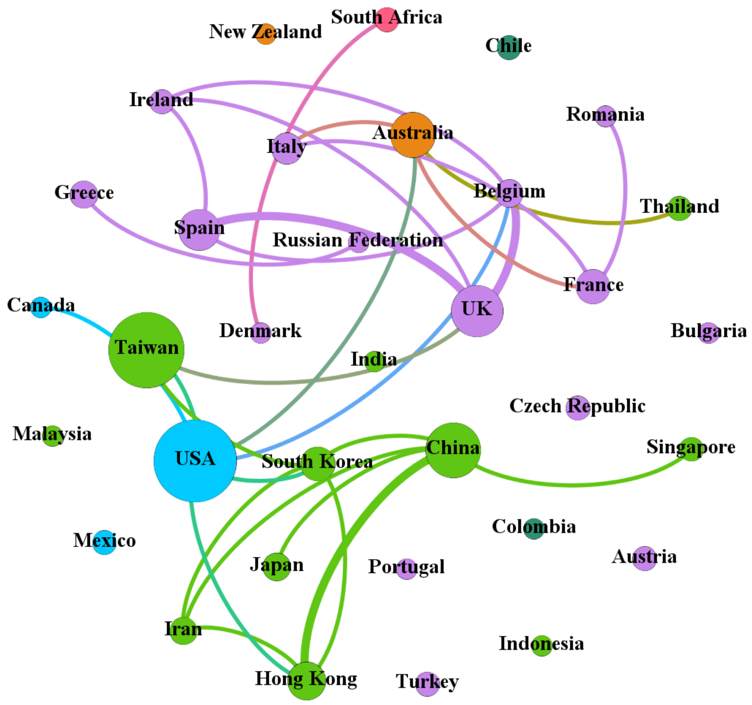


Figure 6. Collaborations among countries/regions

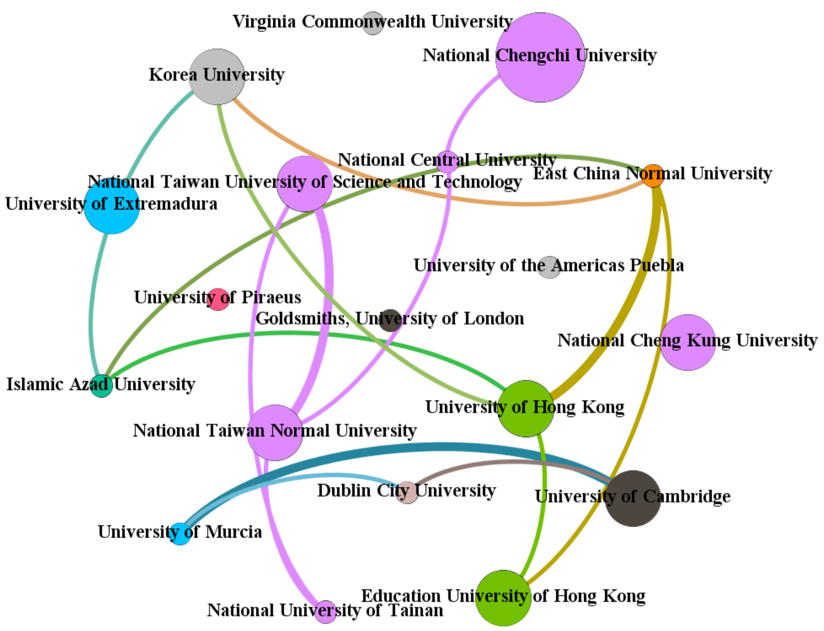


Figure 7. Collaborations among the top 19 prolific institutions

3.4. Results of STM

Figure 8 presents the STM results. The two most popular topics were Mobile-assisted PLL and Anxiety and PLL. According to the Mann-Kendall test results, Personalized grammar learning and Personalized recommendation system for language learning had received significantly increasing research interest.

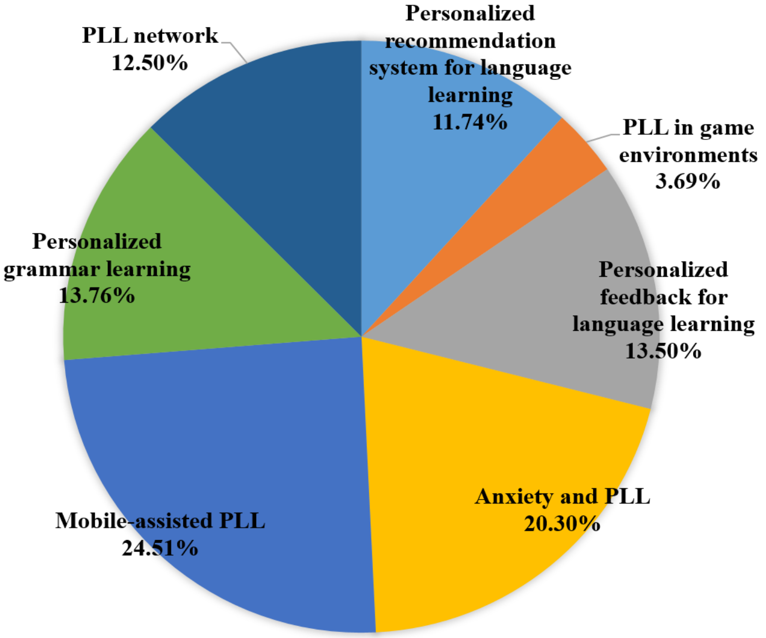


Figure 8. Identified topics with suggested labels and topic proportions

The topic distributions of the top countries/regions and institutions listed in Figures 4 and 5 are visualized in Figures 9 and 10, from which we could see to which research issues each contributor had devoted. For example, Hong Kong and South Korea were interested in Anxiety and PLL. Institutions from Taiwan (e.g., National Central University) devoted much to Mobile-assisted PLL. Such analyses can help countries/regions and institutions identify current and potential scientific strengths and collaborators in PLL research.

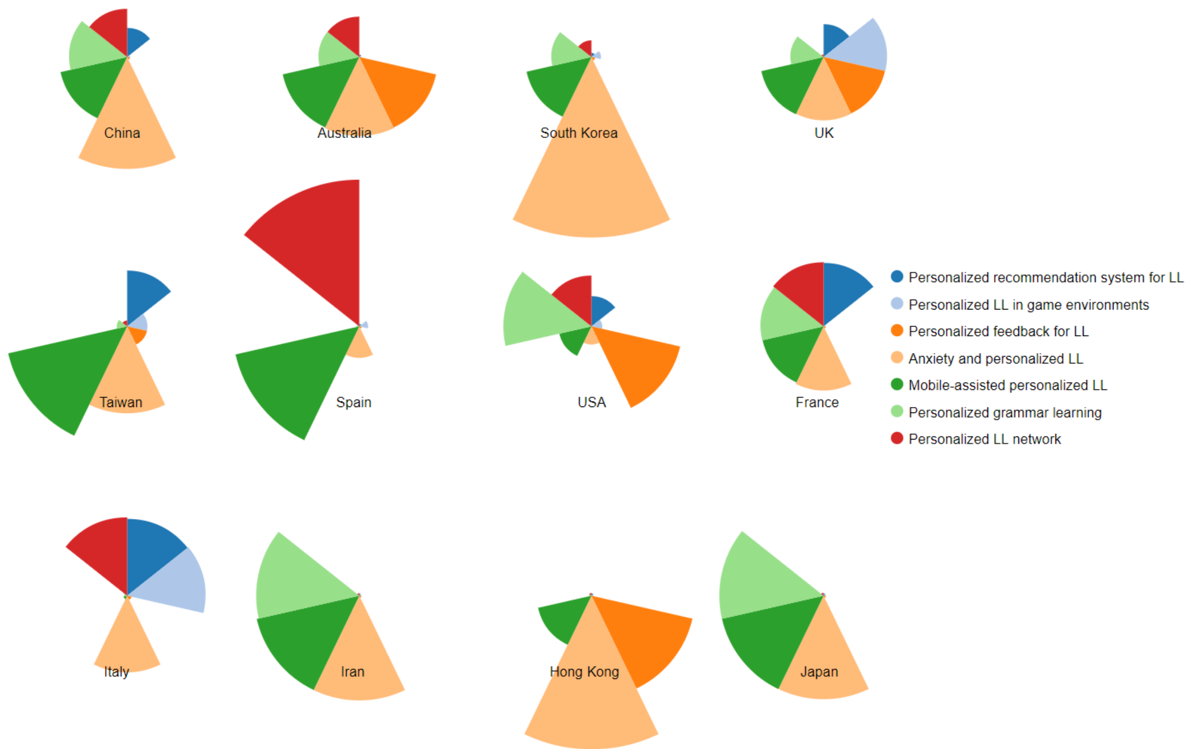


Figure 9. Topic distributions of top countries/regions

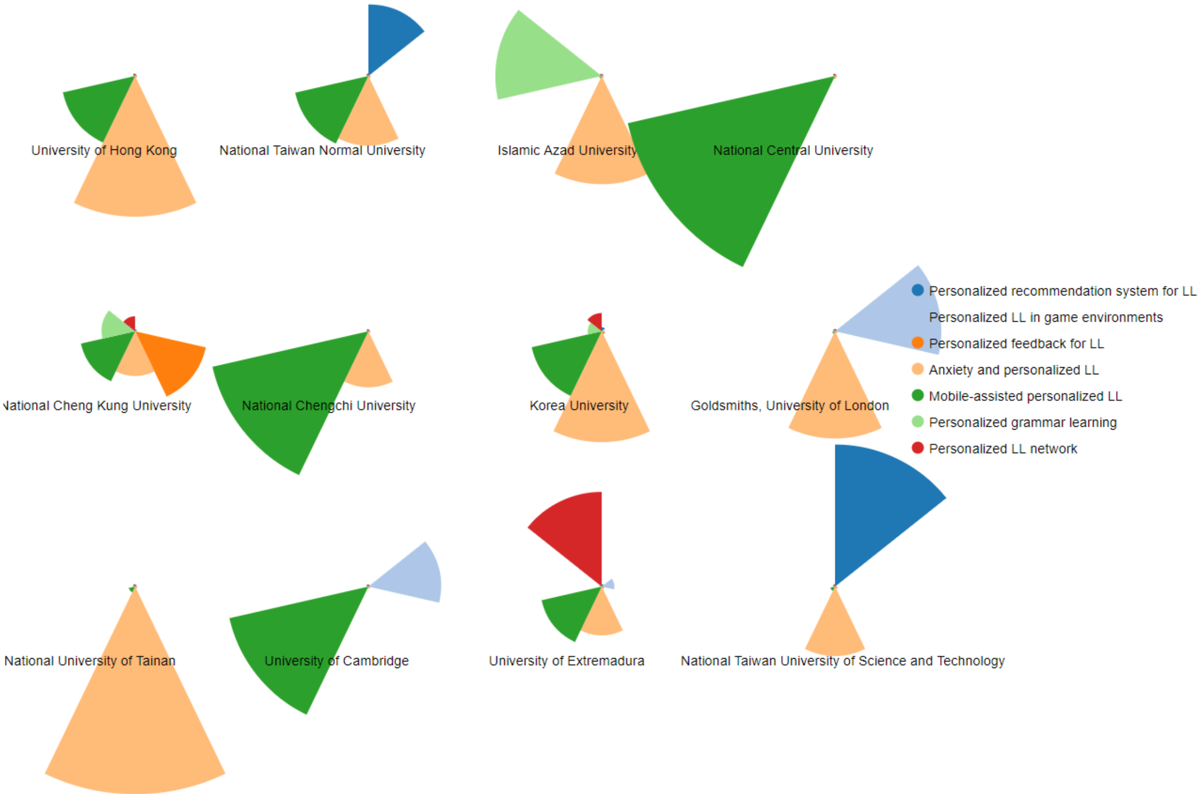


Figure 10. Topic distributions of top institutions

3.5. Results of evolution analysis

Figure 11 shows the evolution of the major phrases used in PLL studies. For a clear presentation, only phrases appearing in more than three studies were considered. Figure 12 shows the emerging phrases in the recent five years. From a technological perspective, very limited technologies (e.g., personal digital assistant) were adopted before 2010. However, due to technological advancement, both the types and applications of innovative technologies (e.g., social media, web 2.0, and computer games) had increased in recent years. For example, mobile devices have become popular since 2015. Intelligent tutoring systems (ITSs) and digital games appeared during 2010–2014 and gained increasing interest since then. There was also a trend in applying LA and AI techniques (e.g., NLP and support vector machines (SVM). From an educational perspective, research enthusiasm about providing personalized feedback increased. Moreover, the realization of PLL based on learner profiles gained increasing attention since the period 2010–2014. Additionally, issues concerning collaboration in PLL started to receive attention in the last few years.

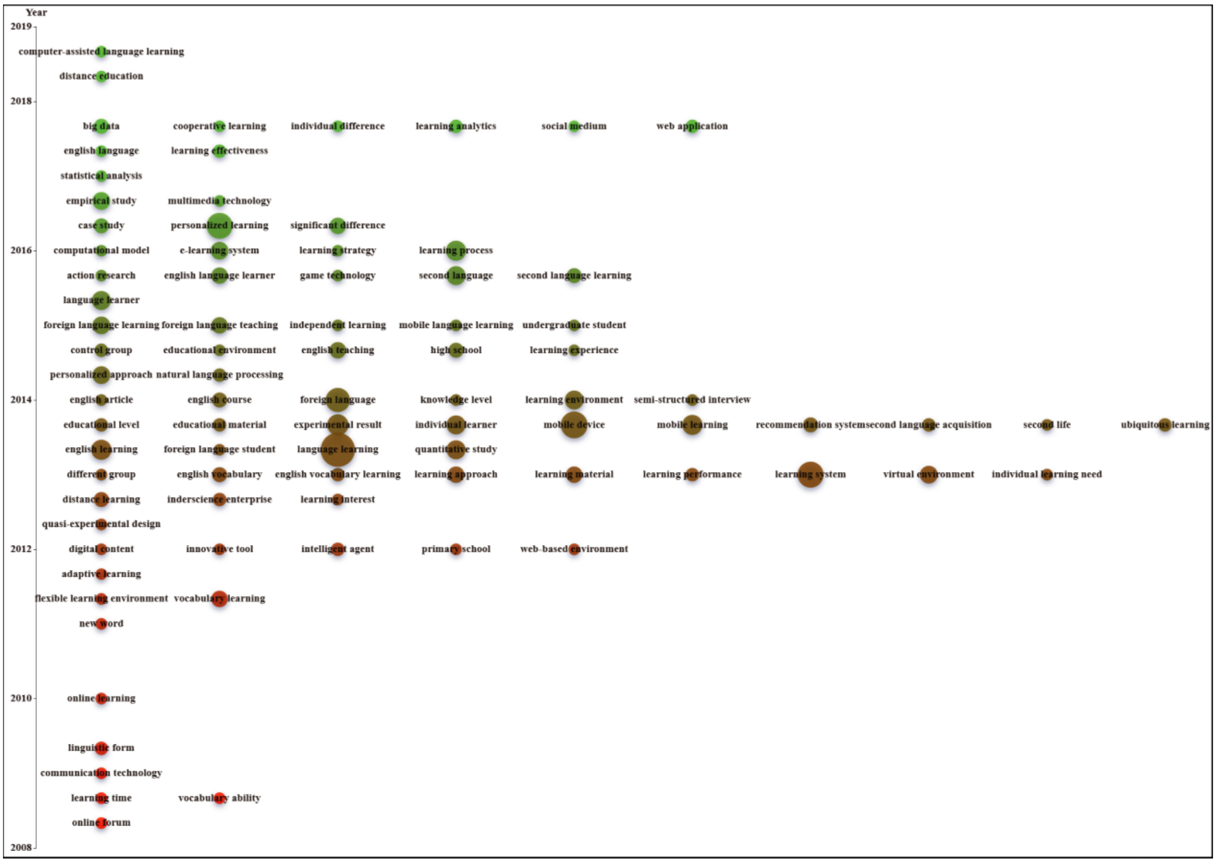


Figure 11. Evolution of major research issues

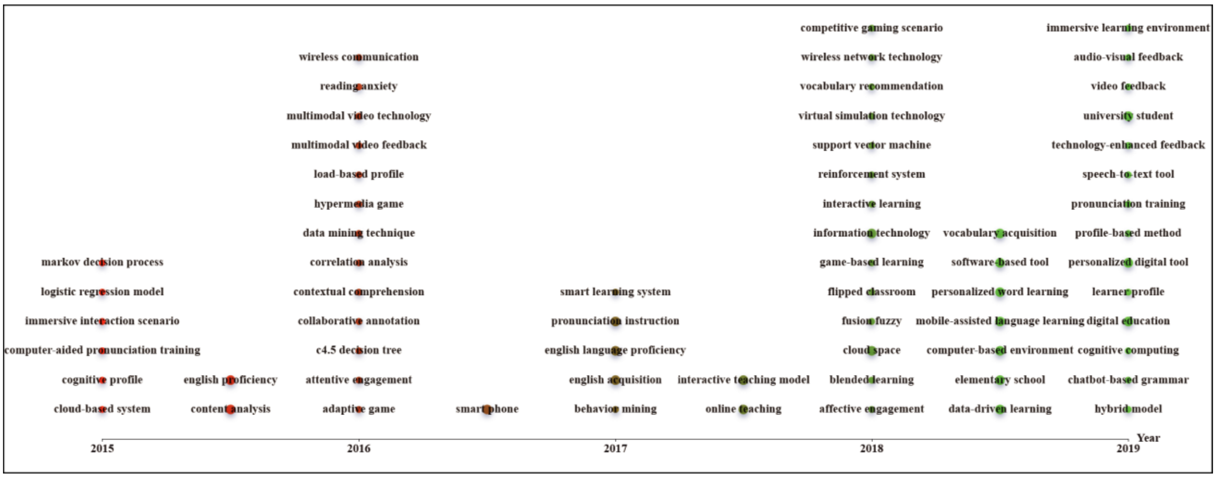


Figure 12. Emerging issues in recent five years

4. Discussions

This study presented a comprehensive overview of PLL research using topic modeling and bibliometrics. The overall increase in academic articles implies that PLL is an increasingly active field with a continuously expanding research community. PLL research enjoys great popularity among interdisciplinary journals that bridge education and technology. The close association between PLL and technology use is also demonstrated by topic and phrase analyses. Countries/regions and institutions (e.g., Taiwan, the USA, and the UK) with large numbers of international collaborations showed better performance and fast development, indicating that international collaboration plays an important role in PLL research to embrace the affordances and face the challenges. Furthermore, the close regional/institutional collaborations are noteworthy, whereas the cross-regional/institutional collaborations should be enhanced. Additionally, institutions with close collaborations tended to show more similarities in research foci. Findings regarding phrase and topic analyses provide insights into future directions for PLL research, as elaborated in the following sub-sections.

4.1. Personalized recommendations in ITSs

The popularity of personalized recommendations in ITSs is associated with its data-driven feature and its ability to reduce the burden of information overload. A data-driven recommendation strategy in ITSs can automatically and dynamically schedule the learning sequences during the learning processes based on learners' performance to further recommend appropriate materials and activities to optimize learning outcomes. In Xie et al. (2015), a recommendation strategy was conducted based on task diversity, word coverage, and context familiarity criteria to determine the subsequent task types, target words, and learning contexts. If a user had completed two reading comprehension tasks that focused on the same target knowledge, then based on task diversity criteria, a cloze task related to the knowledge would be recommended. The system could also rearrange the types and sequence of learning tasks automatically and dynamically based on the learner profiles and learning performance during the learning process. Moreover, a data-driven recommendation strategy in ITSs can balance the suitable recommendations to current knowledge and new learning trajectory exploration. However, optimal recommendations based on dynamic programming is computationally intensive and sometimes infeasible in personalized learning systems. To resolve this, Tang et al. (2019) optimized recommendation systems for adaptive learning through reinforcement learning by iteratively alternating between collecting learners' data using a strategy to choose potential learning materials and improving the strategy using the collected data. Specifically, such a strategy worked by dynamically utilizing current information, for example, knowledge about learning model based on previous students' learning trajectories and the learner information through assessment.

Furthermore, a personalized recommendation strategy in ITSs can reduce the information overload burden through appropriate learning material recommendations to learners based on their interest to provide the "right" information at the "right" time and in the "right" way (Xie et al., 2019). Faced with tasks of different types and varying difficulty levels, few learners know how to make appropriate learning plans due to information overload. According to constructivism and the input hypothesis (Krashen, 1989), it is necessary to recommend tasks with suitable difficulty levels according to learners' prior knowledge levels. Understanding learners' prior knowledge plays a crucial role in optimizing instructional approaches and paces and providing personalized material recommendations that meet learner needs. Such a task could be resolved by analyzing learner profiles or learning portfolios (Xie et al., 2015). In a personalized vocabulary learning system proposed by Xie et al. (2015), load-based learner profiles were constructed to examine and evaluate the involvement load of various tasks and words based on learning logs. The profiles could help effectively optimize recommendations of task types and target words for individual learners.

Our research found that most extant studies concerning personalized recommendations in ITSs are related to reading and vocabulary learning, whereas little research is conducted on personalized grammar learning recommendations. Current grammar learning applications primarily recommend learning materials and exercises according to grammar topics. However, learners are more interested in exercises with similar grammatical structures and usage, particularly those that they have not fully mastered. Thus, future research may consider investigating personalized grammar question recommendations based on learner preferences and needs.

4.2. Personalized feedback and assessment

Technology-driven methods (e.g., automatically analyzing learners' answers, understanding their progress, and adjusting tasks) can enhance learning with the provision of formative assessment and personalized feedback. Our results showed that there is an increase in affective engagement in feedback provided by personalized screencasts and multimodal videos. According to Harper et al. (2018), personalized screencasts for written assignment feedback enhanced students' sense of instructor presence and promoted communication. This might be because hearing the instructor's voice explaining solutions or recommendations for improvement onscreen enabled an increase in students' affective engagement. In line with multimedia learning theory, personalized audio-visual feedback with multimodal formats, conversational tone, and verbal explanations promotes successful engagement with feedback, especially for those with lower language proficiency. Video technologies could significantly enhance computer-mediated communication for language learners by allowing personal appearance and individual language variations, thus promoting interaction and fostering personalized learning and attentive engagement. From the perspective of interaction, personalized video feedback helps build connections with the unconscious mind and emotions, thus creating a platform for socialization between instructors and students and among peers (Hung, 2016). Such socialization benefits learners in providing ideas for peer feedback and learning from their peers' feedback. According to the sociocultural theory, such a scaffolding process enables language knowledge and skill extension. Moreover, according to communities of practice, because videos enable language learners to visualize contexts, body languages, and facial expressions and artifacts, they understand better about their learning and attach more attention to the feedback provision process. Additionally, the contribution of the personalized videos to the sense of personalization can be enhanced by incorporating more extensive use of personal pronouns, hedges, and praise to create a less-distant discourse stance and stronger interpersonal feel.

4.3. Personalized context-aware ubiquitous language learning

Personalized context-aware ubiquitous language learning involves providing personalized and adaptive language learning activities based on learner locations, learning/leisure time, and personal abilities. In Chen et al. (2019), an interactive geographic map (iMap) system was developed to provide learners with a personalized learning environment to facilitate their context-aware ubiquitous learning through detecting locations, considering realistic scenarios, and recommending learning materials accordingly. Learners could navigate within iMap to learn theme-related topics without physical attendance. Also, with iMap, learners could learn about situational dialogues through contextualized dialogue video watching, grasp important vocabularies and phrases, and practice applying knowledge in context to gain mastery. Chen and Li (2010) also developed a personalized context-aware ubiquitous learning system based on learner location as detected by wireless positioning technologies to facilitate situational English vocabulary learning. In such situational learning, social, cultural, and life contexts are used to promote learning in meaningful situations. With the rapid development of location intelligence and wireless technologies, particularly WLAN, which can provide precise location information, personalized context-aware ubiquitous learning can be realized through recommendation mechanism to select language learning materials that associate with learner location from database, so that students experience active interactions with the real world and apply authentic and social knowledge to their surroundings.

4.4. Mobile chatbots for PLL

Mobile chatbots for PLL can be interpreted as integrating chatbots into mobile applications to scale up personalization in language learning by providing services at the learners' convenience with unique learning and human-like interactive learning experiences. Haristiani and Danuwijaya (2019) developed a chatbot-based grammar dictionary application, Gengobot, and integrated it into LINE. When a user sent a grammar item in his/her second language in the Gengobot-integrated LINE, the app would automatically interpret the message, process its intent, and respond with text to provide the meaning of the grammar item, usage patterns, examples, and translations of the examples in the user's first language. In this way, Gengobot enabled users to adjust their learning pace, meet personal needs and preferences, thus supporting PLL. Also, when users communicated with foreign speakers and needed grammar information, they could directly access Gengobot without leaving the chat or exiting LINE. This is different from using a separate application where she or he would need to leave the chat, open the dictionary application, and then return to chat, which is comparatively inconvenient. Moreover, Pham et al. (2018) proposed a mobile English learning application with an integrated chatbot as a virtual personal assistant that automatically responded to casual conversations (e.g., greetings, courtesy, and emotions), recommended personalized learning content (e.g., quizzes, vocabulary or grammar lessons) according to user requests, provided personalized scaffolding or explanations for problem-solving (e.g., helping users to show the term for a definition in a flashcard), and reminded target users to review the previously learned content. This interaction between human beings and chatbots provides a personalized experience for foreign language learning. The social connectivity of mobiles also benefits learners in terms of interactive and personalized learning, particularly when it is combined with chatbots and social media. Nevertheless, our results indicated a lack of critical attention on the development of chatbots as personal assistants to support PLL. Thus, we recommend that more studies focusing on this direction should be conducted.

4.5. Personalized content generation for game-based language learning

Personalized digital game-based language learning can enhance individual learners' professionalism development and command of various language skills via personalized recommendations and tailored advices that are in accordance with their learning needs, preferences, and styles (Hooshyar et al., 2018). Hooshyar et al. (2018) also argued that a data-driven procedural content generation (PCG) method could automatically generate adaptive contents that are tailored to individual learners' language proficiency levels in a game, DLLgame, for children's early reading skill development. In DLLgame, the PCG approach first generated the total of domain-specific contents like learning objectives, materials, and instances. Then, SVM was trained based on data collected from DLLgame to be further added to a genetic algorithm-based content generator to assess content fitness. Subsequently, contents targeting the intended learning outcomes and suiting the players' capabilities were generated and relayed to DLLgame. One specific application of the proposed PCG in DLLgame was to automatically enhance letter recognition through gameplay by identifying and selecting the correct grapheme from various floating letters on the screen. Though given graphemes were presented alphabetically for all players, object pictures for each grapheme were generated by the data-driven PCG strategy in accordance with an individual player's knowledge strengths or deficiencies, thus facilitating the personalized learning experience and enhancing learning performance. In a personalized Portuguese language learning game (Pereira et al., 2012) for vocabulary acquisition, cognitive-based personalization was implemented to continuously assess learners' skill levels and match suitable learning tasks accordingly. The personalization and adaptability of the game could be improved by automatically defining learning levels, tasks, and objectives rather than being predefined by players. Moreover, collaborative game-based learning with interaction and communication features could support social interactivity during gameplay and allow learners to learn along with and from one another, thus enhancing their foreign language learning. Additionally, in collaborative PLL games, instructions, dialogues, and interfaces are presented in a second language, and learners are encouraged to speak or write in a second language to communicate with their collaborators, thus providing intense and meaningful practices of listening and reading.

4.6. AI for personalized diagnosis and adaptation

Moving towards the next generation of PLL environments requires intelligent approaches powered by LA based on powerful AI algorithms that can adapt to individual learning needs for personalized diagnosis and trace of their learning progress (Pérez-Paredes et al., 2018). The rising pervasiveness of AI applications with the ability to predict and adapt based on massive language learner data indicates a promising future of personalized language education. For example, three-dimensional (3D) face recognition techniques are commonly combined with automatic speech recognition (ASR) to provide PLL experience based on individual characteristics. Ming et al. (2013) developed a Mandarin edutainment application for learning Mandarin in an immersive and interactive virtual environment. Specifically, 3D face recognition was used for learner detection, based on which learning materials were adjusted in accordance with the learner's preference and emotional states. Speech recognition interface identified and assessed the spoken content of foreign learners and presented the recognition results onscreen. The proposed system with open, shared, and interactive properties significantly enhanced learner communication ability and created a real-life social community responding to the learners' speech, intent, gesture, and behaviors. Moreover, fuzzy fusion algorithms have demonstrated effectiveness in refining different native speakers' pronunciation characteristics, optimizing personalized evaluation, and improving user feedback adoption rates.

An individual learner's pronunciation error patterns can be detected in an unsupervised manner by using deep learning technologies such as convolutional neural networks (Lee, 2016) which analyzes the acoustic similarity between speech segments from the learner and accommodates variations in pronunciation patterns across students to provide personalized diagnosis and feedback. To diagnose language learning anxiety, Chen et al. (2016) showed the effectiveness of the C4.5 decision tree in facilitating the diagnosis, prediction, and reduction of reading anxiety based on reading annotation behaviors to provide adaptive reading assistance to learners with different levels of learning anxiety. Specifically, the personalized reading anxiety prediction model instantly predicted the reading anxiety of individual learners according to their reading annotation behaviors and interactions with their peers during collaborative reading annotation activities and meanwhile identified the likely reasons for the anxiety using the fired prediction rules determined by a decision tree. The prediction results were displayed on the online tutors' interface, with which the tutors could provide personalized assistance to reduce learners' reading anxiety. Additionally, in a system for automated tracking of learner responses to instructor feedback in draft revision (Cheng et al., 2017), NLP was used to match teacher feedback against syntactic rules and semantic words extracted from annotated data. Specifically, on the syntactic side, every sentence was processed with a part-of-speech tagger to assign parts of speech to each word, while on the semantic side, a word-by-type matrix was constructed based on latent semantic analysis. As the system highlighted gaps between what learners should revise and what they actually revised, students found that it helped them engage in the cognitive process of revisions and reflect critically on gap-bridging, thus promoting effective revisions and high-quality writings. The personalization of the system can be further enhanced by providing personalized suggestions/links to relevant learning resources using collaborative filtering. Based on our results, there is still a lack of critical attention paid to AI techniques, particularly deep learning algorithms.

We thus suggest that scholars keep up with the latest trends in AI by exploring its potential and effectiveness for PLL.

4.7. Personalized LA dashboards

Personalized LA dashboard is a visualized and intuitive display of data to support improvements in personalized learning and performance. In Gelan et al. (2018), learning dashboards were implemented in a Business French course to intuitively visualize online learning behavior to learners and their tutors, provide recommendations inspired by former successful learners, and share the feedback with learners and tutors. Based on the insights obtained from learning visualization through personalized dashboards, learners are provided with personalized recommendations of effective learning strategies, suitable learning resources, and personalized learning pathways to improve learning performance. The connection between personalized learning and LA is significant. Specifically, personalized learning aims at building a profile of each learner's strengths, weaknesses, and learning pace, similar to LA-driven monitoring of learner performance, finding patterns to predict their performance, and customizing instructional support accordingly. Although LA is still in its infancy in implementation and experimentation, it does show great room in its potential use to facilitate PLL, with affordances emerging in data collection and processing, data analysis and interpretation, pattern detection, and learning visualization. LA contributes to PLL research and practices in many ways, for example, allowing for more significant insights into self-directed learning, collecting and analyzing sizable learner data, and promoting personalizing learning through learning dashboards.

4.8. Personalized practice in data-driven learning

Personalized practice in data-driven learning (DDL) is a corpus- and concordance-driven approach for language learning where students use retrieval software to discover rules and draw conclusions by observing and analyzing sizable real corpus and to grasp a grammatical structure or word usage through real-time practice. Particularly, the combination of DDL and mobile devices allows learners to obtain personalized and instant feedback on their own production for practice. For example, a DDL-based mobile app (Pérez-Paredes et al., 2018), TELL-OP, with personalized access features to various corpus-based open educational resources, could effectively support on-the-go PLL due to the instant and personalized assessment via multiple NLP tools and personalized feedback for learners to improve their texts with convenient access to context-sensitive information from various monolingual and collocation dictionaries. DDL emphasizes the active role of learners and their self-directed learning ability to explore language knowledge in accordance with their needs to constantly introspect and induce language rules, which adapts well to the individualization and personalization trends in computer-assisted language learning. Through DDL, the authentic data with various inductive and deductive language learning chances helps individuals get familiar with target language communication and deepen their acquisition of the target knowledge.

4.9. Challenges of PLL

4.9.1. Data policy

While it is common that educational institutions need to store student data to provide an optimal personalized learning experience since such data are critical for automation and personalization, it is important to pay attention to data privacy. Specifically, clear declarations about who can access student learning data, what data they can access, how the data is protected, and what laws protect student data should be provided. Institutions should be transparent about student data privacy practices to dispel misperceptions about student data use and allay concerns (see https://ies.ed.gov/pubsearch/pubsinfo.asp?pubid=NFES2019160).

4.9.2. AI model training

How to choose suitable algorithms is important as each algorithm has advantages and disadvantages. For example, SVM algorithms might show poor performance for sizable data mining since the training of SVMs depends highly on data size. Comparatively, although deep learning algorithms generally show higher performance, their comparatively higher complexity and requirement of expensive computational devices like high-ended GPUs are potential issues regarding their utilization.

4.9.3. Challenges for PLL instructors

Given the learner variability, instructors are challenged to differentiate lessons and to individualize learning for each student while simultaneously keeping the overall level of expectation and rigor in classrooms high. This can be extremely hard since there may be many differences (e.g., personalities, lifestyles, and backgrounds) among students that do not directly impact academic learning. Moreover, since most instructors are not taught about personalized instruction or how to incorporate it into their teaching practice, they usually find it challenging to shift from knowledge imparters to facilitators of learning. Technologies can help instructors overcome obstacles to implementing differentiated or personalized instruction by enabling them to address students' needs in numerous ways (e.g., through content input, learning activities, and opportunities to demonstrate comprehension) (Schmid & Petko, 2019) and to bridge the relationships with students who generally have a predisposition for using tech seamlessly.

4.10. Pedagogical implications for the application of technologies to facilitate PLL

Our findings provide implications for instructors conducting precision language education or are thinking of doing so. To personalize students' learning of pronunciation, listening, and speaking, instructors can take advantage of chatbots, online chat rooms, and VR-based dialogue games. User-friendly mobile phones and chatbot-based applications are effective supplementary tools for regular language curricula to realize language learners' personalized listening and pronunciation practice with voice messages. In an online chat room, students convert their input into the intake in situations where language serves as a communication tool instead of the focus of such interaction. Moreover, VR games can be integrated into pronunciation recognition and assessment systems embedded with ASR to construct a real-life social community responding to the learners' speech, intent, gestures, and behaviors. Additionally, collaborative game-based learning with interaction and communication features can support social interactivity during gameplay and allow students to communicate and interact with one another, thus enhancing their communication and speaking skills.

To recommend personalized materials, we highlight the effectiveness of cluster-specific recommendations, data-driven PCG, similarity-based approach, fuzzy inference mechanisms, context-aware recommendations, and recommendations based on learner emotion. Cluster-specific recommendations help deliver suitable and personalized content to students based on their preferences, learning styles, and learning patterns detected by cluster analysis, sequence analysis, and association rule mining. A data-driven PCG method improves individuals' performance-based gains by providing adaptive content suited to various proficiencies, capabilities, and performance targets of individual learners. Similarity-based recommendation computes the similarity between question query and database questions to generate suitable questions for individual learners. The use of fuzzy inferences and personal memory cycle updates has the potential to find materials best suited for both a learner's ability and her/his need by implicitly modifying memory cycles of content learned before. Moreover, back-propagation neural networks estimate learner location to further recommend appropriate language learning tasks based on context-awareness information to individual learners. Additionally, the personalization of learning material and instructional strategy recommendations can be optimized and adjusted according to individual learners' emotional states identified by face recognition, decision trees, and sentiment analysis techniques. To determine the PLL materials for individual language learners, the integration of students' domain knowledge, working memory capacity, learner profiles, and learning behaviors are key parameters. Nevertheless, comprehensive theoretical frameworks are essential for the optimal design of intelligent PLL systems (Zou & Xie, 2018). For example, in Xie et al. (2015), the involvement load hypothesis was used for incidental word learning task evaluation.

To personalize feedback provision for language learners, instructors are advised to transform feedback from uni-modality to multi-modality and to move the traditional written feedback to video-driven feedback to promote learner engagement. Furthermore, DDL enables learners to receive personalized and immediate feedback right after the completion of learning tasks. Additionally, fuzzy logic supports misconception detection and feedback provision to promote adaptive educational experience by automatically modeling students' learning and forgetting processes using personalized fuzzy inference.

4.11. Future directions for PLL research

Based on our investigation, PLL studies have shown increased use of innovative technologies, particularly in recent years, including multimodal videos, digital games, mobile devices, ITSs, and chatbot, which are effective in providing PLL experience. Additionally, virtual/augmented reality, with advantages in supporting collaborative and immersive learning, can be incorporated in PLL systems to cultivate higher-order thinking skills and communication.

Concerning personalized recommendations, PLL studies focused mainly on providing personalized learning materials. However, research on how to provide the optimal or most suitable recommended materials that match individual needs is lacking, particularly by using ranking-based algorithms to calculate the matching and suitability of each material to target learners. Another potential direction is to involve user decision/opinion to help optimize recommendations by integrating content authoring tools for learners and educators to adaptively add, edit, rate, and evaluate recommended materials, and to make annotations to customize their own recommendation criteria, thus generating more adaptive recommended materials and bringing more personalized learning experience. Additionally, most PLL recommendations were generated based on the learner portfolio and behavioral data analysis. As online learning community activities increase exponentially, social interactions and social connectivity can be considered by incorporating social influence and interactive information (e.g., social media data) between individuals and groups of learners to enhance PLL recommendations.

Concerning emotion recognition, only one study focused on reading anxiety prediction. To further this topic, the development of a merged decision tree considering different types of reading materials to automatically predict the reading anxiety levels of individual learners can be considered. In addition to anxiety, the automatic recognition of learners' overall sentiment in a dynamic and real-time mode can be proposed to further adjust the intelligent learning environments based on learners' emotional states.

In terms of personalized feedback, multimodal videos and screencasting have demonstrated effective for a particular research population. Further study may consider comparing these technologies' affordances in personalized instruction and feedback at different levels of proficiency, or with different languages, learning styles, or skill areas such as the teaching of pronunciation or grammar.

Additionally, we highlight the necessity to keep up with the latest trend in AI (Chen et al., 2020b; Chen et al., 2020c), particularly deep learning (e.g., long short-term memory and generative adversarial network) that have been proven effective in many fields but are currently seldom considered in PLL research. For example, a generative adversarial network has the potential to recommend reading/writing materials of different styles and transform the materials from one style to another based on learners' requirements and needs. Long short-term memory can facilitate PLL in terms of auto-completion and grammar checking, automatic essay scoring, and automatic speech grading when combined with ASR techniques. AI, coupled with deep learning and NLP, has the potential to realize a higher level of personalization by integrating more sophisticated applications that are able to adapt, learn, and predict with ultimate autonomy. In sum, attention should reach beyond computer-related technologies to the latest AI technological trends and their applications in language education to construct language knowledge and develop critical thinking, thus promoting language learning achievement in a personalized way. Particularly, inter-department collaboration among system developers, language experts, and practitioners should be enhanced to provide more effective and easier-to-use PLL systems representing true integration of language pedagogy, language practice, and technology.

4.12. Limitations

There are several limitations in this study. Firstly, our analysis was based on records retrieved from WoS and Scopus. Although WoS and Scopus are multidisciplinary databases of academic output and are commonly adopted for literature reviews, there might still be PLL articles that were not included in the two databases. Furthermore, this study used merely English language research articles. However, as PLL is being explored worldwide, publications written in other languages should also be considered in future research. Additionally, as mentioned in section 2, in data screening, most of the articles were excluded due to irrelevance to personalized learning (N = 40) or language learning (N = 314). However, they were retrieved initially due to the share of search strings by studies from different research fields. Future studies may consider optimizing the search strategy by using a more context-specific search query.

5. Conclusions and significance

Our study was the first-in-depth to track current advances in PLL research using STM and knowledge mapping. Such timely work is needed with PLL attracting increasing attention from academia, although the total number of studies is small at the current stage. A large proportion of PLL studies focused on providing personalized recommendations on learning materials and tasks and personalized feedback, while issues such as emotion detection, cognitive loads, and higher order thinking skills, receiving relatively less attention. Technologies play an important role in facilitating PLL, with various applications (e.g., mobile devices, digital games, ITSs, multimodal videos, and wireless technologies), analytical techniques (e.g., LA, PCG, and AI), and learning strategy (e.g., DDL) being increasingly adopted with positive effectiveness being reported. Language learners experiencing personalized learning generally showed improvement in learning gains and engagement, satisfaction with the personalized learning experience, and an increase in self-efficacy and confidence. Personalized language learners were also viewed as having higher learning motivation and a positive attitude toward language learning, together with higher acceptance toward technologies involved in the learning process. In sum, our study indicates that (1) multimodal videos promote effective personalized feedback; (2) personalized context-aware ubiquitous language learning enables active interaction with the real world by applying authentic and social knowledge to their surroundings; (3) mobile chatbots with ASR provide human-like interactive learning experiences to practice speaking and pronunciation; (4) collaborative game-based learning with customized gameplay path and interaction and communication features supports social interactivity and learning of various language skills; (5) AI promotes effective outcome prediction and instruction adaptation for individuals based on massive learner data; (6) LA dashboards facilitate personalized recommendations through learning data visualization; and (7) DDL allows personalized and immediate feedback in real-time practice.

Acknowledgement

This research was supported by the Faculty Research Fund (102041) and the Lam Woo Research Fund (LWI20011) of Lingnan University, Hong Kong, the One-off Special Fund from Central and Faculty Fund in Support of Research from 2019/20 to 2021/22 (MIT02/19-20), the Research Cluster Fund (RG 78/2019-2020R), and the Interdisciplinary Research Scheme of the Dean's Research Fund 2019-20 (FLASS/DRF/IDS-2) of The Education University of Hong Kong, Hong Kong.

References

Chen, C. M., & Li, Y. L. (2010). Personalised context-aware ubiquitous learning system for supporting effective English vocabulary learning. *Interactive Learning Environments*, *18*(4), 341–364.

Chen, C. M., Wang, J. Y., Chen, Y.-T., & Wu, J. H. (2016). Forecasting reading anxiety for promoting English-language reading performance based on reading annotation behavior. *Interactive Learning Environments*, *24*(4), 681–705.

Chen, M. P., Wang, L. C., Zou, D., Lin, S. Y., & Xie, H. (2019). Effects of caption and gender on junior high students' EFL learning from iMap-enhanced contextualized learning. *Computers & Education, 140*, 103602. doi:10.1016/j.compedu.2019.103602

Chen, X., Xie, H., Wang, F. L., Liu, Z., Xu, J., & Hao, T. (2018). A Bibliometric analysis of natural language processing in medical research. *BMC Medical Informatics and Decision Making*, *18*(1), 1–14. doi:10.1186/s12911-018-0594-x

Chen, X., Zou, D., Cheng, G., & Xie, H. (2020a). Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A Retrospective of all volumes of computer & education. *Computers & Education*, *151*, 103855. doi:10.1016/j.compedu.2020.103855

Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020b). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *1*, 100002. doi:10.1016/j.caeai.2020.100002

Chen, X., Xie, H., & Hwang, G. J. (2020c). A Multi-Perspective study on artificial intelligence in education: Grants, conferences, journals, software tools, institutions, and researchers. *Computers and Education: Artificial Intelligence*, *1*, 100005. doi:10.1016/j.caeai.2020.100005

Chen, X., Zou, D., & Xie, H. (2020d). Fifty years of British Journal of Educational Technology: A Topic modeling based bibliometric perspective. *British Journal of Educational Technology*, *51*(3), 692–708.

Cheng, K., Chwo, S., Chen, J., Foung, D., Lam, V., & Tom, M. (2017). Automatic classification of teacher feedback and its potential applications for EFL writing. In *Proceedings of the 25th ICCE. New Zealand: Asia-Pacific Society for Computers in Education* (pp. 884–889). Christchurch, New Zealand: Asia-Pacific Society for Computers in Education.

Chrysafiadi, K., & Virvou, M. (2013). Dynamically personalized e-training in computer programming and the language C. *IEEE Transactions on Education*, *56*(4), 385–392.

Connor, C. M., Mazzocco, M. M. M., Kurz, T., Crowe, E. C., Tighe, E. L., Wood, T. S., & Morrison, F. J. (2018). Using assessment to individualize early mathematics instruction. *Journal of School Psychology*, *66*, 97–113. doi:10.1016/j.jsp.2017.04.005

Fang, L., Tuan, L. A., Hui, S. C., & Wu, L. (2018). Personalized question recommendation for English grammar learning. *Expert Systems, 35*(2), e12244. doi:10.1111/exsy.12244

Gelan, A., Fastré, G., Verjans, M., Martin, N., Janssenswillen, G., Creemers, M., Depaire, B., & Thomas, M. (2018). Affordances and limitations of learning analytics for computer-assisted language learning: A Case study of the VITAL project. *Computer Assisted Language Learning*, *31*(3), 294–319. doi:10.1080/09588221.2017.1418382

Hao, T., Chen, X., Li, G., & Yan, J. (2018). A Bibliometric analysis of text mining in medical research. *Soft Computing*, *22*(23), 7875–7892.

Haristiani, N., & Danuwijaya, A. R. I. A. (2019). Gengobot: A Chatbot-based grammar application on mobile instant messaging as language learning medium. *Journal of Engineering Science and Technology*, *14*(6), 3158–3173.

Harper, F., Green, H., & Fernandez-Toro, M. (2018). Using screencasts in the teaching of modern languages: Investigating the use of Jing® in feedback on written assignments. *The Language Learning Journal*, *46*(3), 277–292.

Hooshyar, D., Yousefi, M., Wang, M., & Lim, H. (2018). A Data-driven procedural-content-generation approach for educational games. *Journal of Computer Assisted Learning*, *34*(6), 731–739.

Hung, S.-T. A. (2016). Enhancing feedback provision through multimodal video technology. *Computers & Education*, *98*, 90–101. doi:10.1016/j.compedu.2016.03.009

Ismail, H. M., Harous, S., & Belkhouche, B. (2016). Review of personalized language learning systems. In *2016 12th International Conference on Innovations in Information Technology (IIT)* (pp. 1–6). Al-Ain, United Arab Emirates: IEEE. doi:10.1109/INNOVATIONS.2016.7880051.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The modern language journal, 73*(4), 440–464.

Lee, A. (2016). *Language-independent methods for computer-assisted pronunciation training* (Doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

Lian, A.-P., & Sangarun, P. (2017). Precision language education: A Glimpse into a possible future. *GEMA Online® Journal of Language Studies*, *17*(4), 1–15. doi:10.17576/gema-2017-1704-01

Lu, O., Huang, A., Huang, J., Lin, A., Ogata, H., & Yang, S.J.H. (2018). Applying learning analytics for the early prediction of students' academic performance in blended learning. *Educational Technology & Society*, *21*(2), 220–232.

Ming, Y., Ruan, Q., & Gao, G. (2013). A Mandarin edutainment system integrated virtual learning environments. *Speech Communication*, *55*(1), 71–83.

Pereira, A. B. C., Junior, G. S., Monteiro, D. C., Barros, E. S., Costa, H. P., Nascimento, P. A. A., Marques, L. B., de Souza, D. G., Salgado, F. M., & Bessa, R. Q. (2012). A AIED game to help children with learning disabilities in literacy in the Portuguese language. In *2012 Brazilian Symposium on Games and Digital Entertainment* (pp. 134–143). Brasilia, Brazil: IEEE. Retrieved from https://www.sbgames.org/sbgames2012/proceedings/papers/computacao/comp-full_17.pdf

Pérez-Paredes, P., Ordoñana Guillamón, C., & Aguado Jiménez, P. (2018). Language teachers' perceptions on the use of OER language processing technologies in MALL. *Computer Assisted Language Learning*, *31*(5–6), 522–545.

Pham, X. L., Pham, T., Nguyen, Q. M., Nguyen, T. H., & Cao, T. T. H. (2018). Chatbot as an intelligent personal assistant for mobile language learning. In *Proceedings of the 2018 2nd International Conference on Education and E-Learning* (pp. 16–21). Bali, Indonesia: ACM. doi:10.1145/3291078.3291115

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, *58*(4), 1064–1082.

Schmid, R., & Petko, D. (2019). Does the use of educational technology in personalized learning environments correlate with self-reported digital skills and beliefs of secondary-school students? *Computers & Education*, *136*, 75–86.

Tang, X., Chen, Y., Li, X., Liu, J., & Ying, Z. (2019). A Reinforcement learning approach to personalized learning recommendation systems. *British Journal of Mathematical and Statistical Psychology*, *72*(1), 108–135.

US Department of Education. (2017). *Reimagining the Role of Technology in Education: 2017 National Education Technology Plan Update*. Retrieved from https://tech.ed.gov/files/2017/01/NETP17.pdf

Wu, T.-T., Huang, Y.-M., Chao, H.-C., & Park, J. H. (2014). Personalized English reading sequencing based on learning portfolio analysis. *Information Sciences*, *257*, 248–263. doi:10.1016/j.ins.2011.07.021

Xie, H., Chu, H.-C., Hwang, G.-J., & Wang, C.-C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: A Systematic review of journal publications from 2007 to 2017. *Computers & Education*, *140*, 103599. doi:10.1016/j.compedu.2019.103599

Xie, H., Zou, D., Lau, R. Y. K., Wang, F. L., & Wong, T.-L. (2015). Generating incidental word-learning tasks via topic-based and load-based profiles. *IEEE Multimedia*, *23*(1), 60–70.

Yang, S. J. H. (2019). Precision education: New challenges for AI in education. In *Proceedings of the 27th International Conference on Computers in Education (ICCE)* (Keynote speech). Kenting, Taiwan: Asia-Pacific Society for Computers in Education (APSCE). Retrieved from https://youtu.be/VKmUE1HnaroZhang, L., Basham, J. D., & Yang, S. (2020). Understanding the implementation of personalized learning: A Research synthesis. *Educational Research Review*, *31*, 100339. doi:10.1016/j.edurev.2020.100339

Zhang, R., & Zou, D. (2020). Types, purposes, and effectiveness of state-of-the-art technologies for second and foreign language learning. *Computer Assisted Language Learning*, 1–47. doi:10.1080/09588221.2020.1744666

Zou, D., & Xie, H. (2018). Personalized word-learning based on technique feature analysis and learning analytics. *Educational Technology & Society*, *21*(2), 233–244.

Unfolding Students' Online Assignment Submission Behavioral Patterns using Temporal Learning Analytics

Mehmet Kokoç[1], Gökhan Akçapınar[2*] and Mohammad Nehal Hasnine[3]
[1]School of Applied Sciences, Trabzon University, Turkey // [2]Faculty of Education, Hacettepe University, Turkey // [3]Research Center for Computing and Multimedia Studies, Hosei University, Japan //
kokoc@trabzon.edu.tr // gokhana@hacettepe.edu.tr // nehal.hasnine.79@hosei.ac.jp
[*]Corresponding author

ABSTRACT: This study analyzed students' online assignment submission behaviors from the perspectives of temporal learning analytics. This study aimed to model the time-dependent changes in the assignment submission behavior of university students by employing various machine learning methods. Precisely, clustering, Markov Chains, and association rule mining analysis were used to analyze students' assignment submission behaviors in an online learning environment. The results revealed that students displayed similar patterns in terms of assignment submission behavior. Moreover, it was observed that students' assignment submission behavior did not change much across the semester. When these results are analyzed together with the students' academic performance at the end of the semester, it was observed that students' end-of-term academic performance can be predicted from their assignment submission behaviors at the beginning of the semester. Our results, within the scope of precision education, can be used to diagnose and predict students who are not going to submit the next assignments as the semester progresses as well as students who are going to fail at the end of the semester. Therefore, learning analytics interventions can be designed based on these results to prevent possible academic failures. Furthermore, the findings of the study are discussed considering the development of early-warning intervention systems for at-risk students and precision education.

Keywords: Precision education, Temporal learning analytics, Educational data mining, Assignment submission behavior, Learning performance

1. Introduction

A deeper understanding of online learning experiences is required for learning designers and researchers. Studies on theory and practice on how students learn individually or in groups in online environments by analyzing students' trace data have increased in recent years (Yang et al., 2020). In the last decade, learning analytics (LA) studies employing machine learning methods have been carried out to gain actionable insights such as at-risk students' detection, learning outcome assessment, and drop-out detection for improving the teaching quality and learning process. Precision education is known to be a relatively new discipline in higher education that uses the core philosophies of LA and data-driven methods. Precision education is, as addressed by (Yang, 2019), a new challenge for conventional LA, machine learning, and artificial intelligence for solving critical aspects in online education such as spotting at-risk, drop-out, low-engaged students as early as possible by analyzing online learning behaviors (for instance, assignment submission pattern, and engagement with learning materials). Precision education contributes towards maximizing students' online learning experiences and value proposition, and therefore, it uses data from the latest learning technology and integrates student support processes to ensure the highest quality teaching (Wilson & İsmaili, 2019). One of the goals of precision education is to predict students' learning performance by analyzing their online learning behaviors and providing timely intervention for supporting their learning process (Lu et al., 2018). Furthermore, precision education can be leveraged to uncover various critical aspects of education including behavioral, cognitive and emotional.

While precision education emphasizes employing artificial intelligence and other data-driven methods on large-scale datasets collected from technology-enhanced learning environments (i.e., learning management systems, digital textbooks), data about assignment submission behavior can be explored more within the scope of precision education. Students' online assignment submission behavior is a meaningful part of online learning experiences (Akçapınar & Kokoç, 2020) and has a relationship with procrastination (Yang et al., 2020). Students' online assignment submission behavior could reveal information about learning behavior such as how students' behavioral patterns of online assignment submission change over time or relationships between students' online assignment submission behaviors and their learning performance. These insights on learning behavior are crucial for teachers to monitor their students' learning progress, particularly to spot at-risk or inattentive students as early as possible. Therefore, modeling students' online learning behaviors hidden in the learning traces is an important LA contribution for precision education. Thus, modeling students' online assignment submission behavior using temporal analysis techniques can provide important insights into the online learning process and help teachers to plan timely interventions for procrastinators and/or at-risk students for precision education. Furthermore, the temporal aspect of online assignment submission behavior in precision education has much to offer in diagnosing students' learning behavior, however, not been explored much.

As of now, much effort has given to explore learning behavior patterns and predict learning performances based on interaction data, far too little attention has been paid to analyzing temporal and sequential aspects of trace data of students (Chen, Knight, Wise, 2018; Olsen, Sharma, Rummel, & Aleven, 2020). Several studies have focused mainly on aggregated data (e.g., the total number of events) without considering temporal aspects of online learning behaviors (Juhaňák, Zounek, & Rohlíková, 2019). To the best of our knowledge, a limited number of studies detect patterns in students' online assignment submission behaviors using temporal analysis techniques. Considering the importance of temporal analytics in precision education for diagnosing students' learning, behavioral patterns, and learning performance prediction, this study explored students' online assignment submission behavior patterns by using clustering, Markov Chains, and association rule mining analysis. With these analyses, this study aimed to contribute precision education literature to investigate whether these patterns can be used to diagnose and predict at-risk students (e.g., who are not going to submit next assignment and low-performers) as early as possible. Our study addressed the following research questions:

RQ1. What are the students' behavioral patterns of online assignment submission?
RQ2. How do students' behavioral patterns of online assignment submission change over time?
RQ3. What are the association rules between students' online assignment submission behaviors and their learning performance that can be used to predict at-risk students as early as possible?

This paper aims to employ educational data mining methods for precision education to uncover a core focus of precision education, namely, understanding learning behavior while the semester progresses. More precisely, this paper aimed to diagnose and predict at-risk students based on their online assignment submission behavior over time using temporal LA. Students' online submission behavioral data were collected from Moodle and analyzed with regards to- how students' assignment submission behavior changes over the period of time, finds association between their assignment submission and final score, and analyzes the factors affecting students' learning performance at the end of the semester; and visualizes students' assignment submission patterns so that the teacher can get an early insight about the students. Thus, the predictive models and the findings of this study contribute to the core of precision analytics.

2. Background and literature review

2.1. Precision education

Employing artificial intelligence and machine learning techniques in education and psychology has led to significant developments in related fields such as educational intelligence, self-regulated learning, and precision education. Depending on the developments in information and communication technologies, a paradigm shift in learning and teaching has occurred and new pedagogical models have emerged. One of the new educational models considering personalized learning is precision education. Precision education can be defined as a new challenge of applying artificial intelligence, machine learning, and LA for improving teaching quality and learning performance (Yang, 2019). Precision education aims to analyze educational and learner data, predict students' performance and provide timely interventions based on learner profiles for enhancing learning (Lu et al., 2018). For effective learning design in precision education, LA has contributed not only to the dashboards and intervention tools but also as the conceptual frameworks guiding research experiences.

The ultimate goals of using LA are to increase student success and improve students' online learning experience (Pardo & Dawson, 2016). Studies in LA and precision education literature have provided new findings based on multimodal data and actionable knowledge to increase the learning/teaching context's effectiveness. There have been several attempts (e.g., Azcona, Hsiao, & Smeaton, 2019; Tsai et al., 2020) to explore students' interaction and behavioral patterns, to predict students' learning performance based on their online learning behaviors, to develop early-warning systems for at-risk students, to support students and teachers decision-making processes, and to investigate effects of interventions and LA dashboards. The results of the aforementioned studies indicate that LA provides important clues about students' online learning experiences and LA tools offer personalized recommendations to students by visualizing and analyzing their trace data to optimize and improve learning. It is clear that LA and employing educational data mining methods in educational studies contributes to our understanding of learning.

While many studies have been carried out on profiling learners and prediction learning performances based on interaction data, less attention has been paid to analyzing temporal and sequential aspects of trace data of students (Chen, Knight, & Wise, 2018; Juhaňák, Zounek, & Rohlíková, 2019). Rather than modeling the frequency of clicks and interaction of students in an online learning environment, students' learning paths need to be modeled based on time and probability (Cerezo, Sánchez-Santillán, Paule-Ruiz, & Núñez, 2016). Thus, there is an important gap in the relevant field in terms of behavior modeling. To overcome this gap, event logs reflecting students' learning experiences have been modeled using temporal analysis and temporal LA approach (Knight, Wise, & Chen, 2017). The following section is about temporal LA and its implementation in the educational context. In precision education, the diagnosis of online learning behavior patterns for using predictive student modeling is vital to provide students real-time intervention.

2.2. Temporal LA and its role in the educational context

Literature in the educational contexts indicates that both individual and collaborative learning do not happen in one moment (Knight, Wise, & Chen, 2017). In general, learning happens over a period, which is referred to as a process. Temporal characteristics of students' learning data contain valuable insights about the time period or process of occurrence of particular events (Mahzoon et al., 2018). Thus, analyzing time-related data rather than just frequencies gives more information about the learning process (Knight, Wise, & Chen, 2017). The temporal analysis of students' learning data provides a more in-depth insight into individual and collaborative learning processes (Nguyen, Huptych, & Rienties, 2018; Olsen et al., 2020). What makes temporal analysis vital in online and blended learning is that modeling transitions between different students' actions considering temporal changes enhance our understanding of online learning behavioral patterns. Also, temporal analysis supports a more robust prediction model of students' learning performance to make timely interventions for precision education.

In the temporal analysis, various techniques are employed for modeling students' behaviors extracted from their trace data include process mining, sequential pattern mining, Markov chains, and hidden Markov models. While process mining discovers a process model from the students' activity sequences, sequential pattern mining finds the most frequent patterns through a range of action sequences. Markov chains aggregates sequences of students' actions into transition models and hidden Markov models have been used for discovering students' behavioral patterns considering transitions over time (Boroujeni & Dillenbourg, 2019). There is a significant difference between time-series analysis and temporal LA. While time-series analysis typically looks for recurring patterns within a time period for numeric features (Mahzoon et al., 2018), temporal analytics methods help researchers analyze dynamic student data and mode student behaviors over time at different levels of granularity.

There is an increasing trend of temporal analytics methods being used to diagnose students' online learning behavior patterns and predict their learning performance based on temporal data for planning timely interventions (Cheng et al., 2017; Juhaňák, Zounek, & Rohlíková, 2019; Matcha et al., 2019). Previous studies have shown that temporal analytics is beneficial to predict students' learning performance (Papamitsiou & Economides, 2014), to diagnose of learning patterns and behaviors (Boroujeni & Dillenbourg, 2019), to identify at-risk learners (Mahzoon et al., 2018), to detect learning tactics and strategies (Matcha et al., 2019) and to explore the relationship between students' timing of engagement and learning design (Nguyen, Huptych, & Rienties, 2018). While the importance of analyzing students' temporal trace data in online and blended learning has great potential in improving educational practice, applying temporal analytics to student data is less explored in educational research (Chen, Knight, & Wise, 2018; Knight, Wise, & Chen, 2017). To date, the temporal analysis of trace data has been mostly employed in modeling students' online behaviors in the LA field (Juhaňák, Zounek, & Rohlíková, 2019). These studies highlight the critical role of temporal analysis of trace data in diagnosing online learning behaviors and predicting students' further actions. Although temporal analysis has been used to unlock students' online learning behaviors such as quiz-taking, content navigation, e-book reading, and video viewing, few studies have paid attention to exploring online assignment submission behavior patterns. Therefore, in our study, we intended to use the temporal LA method to model students' online learning behavior patterns, specifically students' trace data while engaging in online assignment activities.

2.3. Online assignment submission behaviors

There is an increasing demand for online assignments to assess the learning process and evaluate learning performance. Submission of online assignments is one of the most performed online learning activities by students (Cerezo et al., 2016). In addition, assignment activity is a commonly used LMS component in blended learning environments and fully online courses (Azcona, Hsiao, & Smeaton, 2019). Moreover, several studies have shown that number of submitted online assignments, assignment scores, and interaction with assignments are predictors of students' learning performances (Lu et al., 2018; Zacharis, 2015). According to a study that modeled LMS-generated interaction data, students' interaction with assignments and learning tasks are vital parts of their learning experiences (Kokoç & Altun, 2019). Since online assignments play a meaningful role both in evaluating to what extent students understand the course subjects and practicing a course topic (Tila & Levy, 2020), online assignment submission behavior can have crucial consequences for learning process assessment. Thus, the diagnosis of students' online assignment submission behaviors has been the subject of much attention in the literature. Previous studies indicated that students who uploaded their assignments previous to the submission deadline had been better online learning experiences and higher course performance (Akçapınar & Kokoç, 2020; Paule-Ruiz, Riestra-González, Sánchez-Santillán, & Pérez-Pérez, 2015).

One of the key educational aspects that makes online assignment submission times vital for precision education is the early identification of students with procrastination tendencies (Yang et al., 2020). Students' online assignment submission times have been added to the LA indicators as a proxy measure of academic procrastination for identifying students at risk of failure (Cormack, Eagle, & Davies, 2020). For example, Yang et al. (2020) predicted students' academic performance through submission pattern data reflecting their procrastination behaviors with an accuracy of 97%. Additionally, previous studies showed that delaying online assignment submission as a procrastination behavior resulted in lower grades (Cerezo, Esteban, Sánchez-Santillán, & Núñez, 2017; Cormack, Eagle, & Davies, 2020). This indicates the importance of analyzing online assignment submission behavior to identify at-risk and procrastinator students for precision education.

Previous studies indicated that the late completion of an online assignment was associated with lower academic performances and procrastination tendencies (Cormack, Eagle, & Davies, 2020; Yang et al., 2020). Whereas online assignment submission behavior is essential for the prediction of students' learning performance and understanding their online learning experiences, little is still known about it from temporal LA perspectives. To the best of our knowledge, only one study by Akçapınar and Kokoç (2020) analyzed students' online assignment submission behaviors and found that three clusters emerged based on submission behaviors and most of the students who did not submit the assignment failed in the blended course. Although this study provides valuable results on the assignment submission behavior process, more LA research is needed to expand our understanding of online assignment submission behavior in an online and blended learning environment, especially following temporal analysis and modeling (Azcona, Hsiao, & Smeaton, 2020; Yang et al., 2020). Understanding the process of students' online assignment submission behavior can provide important insights into an effective personalized/adaptive learning environment and help teachers to plan timely interventions for procrastinators and/or at-risk students for precision education. Thus, our study aims to better understand students' online assignment submission transition behaviors by visualizing the patterns and predicting their further assignment behaviors in a blended learning course. We hope that the study sheds some light on online assignment submission behavioral patterns and provides actionable knowledge to design timely interventions for improving learning.

3. Method

In order to answer the research questions, students' assignment submission data were analyzed using state-of-the-art educational data mining techniques including clustering, Markov Chains, and association rule mining. Markov models and clustering and predictive analysis are commonly used in precision education research as they can generate easy-to-understand models to diagnose and predict at-risk students on time by analyzing their behavioral data collected from the educational learning environments (Boroujeni & Dillenbourg, 2019). These methods can also help researchers to understand the transition probabilities of different students' behaviors that can be valuable to plan further interventions to prevent possible academic failures. The employed combined method allows us to obtain interpretable models to understand the students' assignment submission behavior, its relation with the academic performance, and changes that happened over time. The data collection and data analysis processes are explained in detail in the following sections.

3.1. Participants and context

The data were collected from an Operating Systems course offered by a public university in Turkey. A total of sixty-nine students participated in the study. In this course, Moodle was actively used as a part of the lecture delivery together with face-to-face lessons. The students' activities in Moodle can be summarized as following the course resources, participating in the discussions, and doing assignments. The assignments included open-ended questions related to the weekly topics. The purpose of the assignments was to make the students come prepared for the class. Students are given five-six days before the class to complete the assignments. The starting time of the class was set as the deadline for the assignment of the last week. During the semester, 10 assignments were given to the students. In this study, the data related to the assignment given to the students in the 4^{th}, 6^{th}, 8^{th}, and 10^{th} week were analyzed. These assignments are chosen because they are directly related to course objectives. The instructor prepares questions in quizzes to promote students' use of higher-order thinking skills such as remembering, understanding, applying, analyzing, revising, and creating. An example of a question related to the disk scheduling topic is given below. In order to answer this question, the students must know how the disk scheduling algorithms work and apply them to the given context.

Example Question: Let's take an example where the queue has the following requests with cylinder numbers as follows: 90, 198, 27, 112, 16, 104, 69, and 60. Assume the head is initially at cylinder 50. Sort incoming requests according to the SSTF (shortest-seek-time-first) algorithm.

The students submitted their assignments through the Quiz module in Moodle. Among 69 students, 48 students submitted the first assignment, 57 students submitted the second assignment, 50 students submitted the third assignment, and 48 students submitted the fourth assignment. The events that students can perform in the assignment submission process are presented in Table 1. All the activities related to these events were logged in Moodle's database with a time stamp.

Table 1. Activities that the students can perform in the assignment submission process

Event	Description
Assignment viewed	The student viewed the assignment module, saw the assignment description, but did not open the questions.
Attempt started	This is only the case when the student views the assignment for the first time, and this does not happen again on subsequent visits.
Question viewed	The student's displaying each question in the assignment is logged in this way. Displaying the question also means recording the text in the answer field.
Assignment submitted	This happens when the student completes the assignment. The student can submit the assignment once and then cannot change the answers.
Question reviewed	If the student displays the assignment after the deadline, it will be labeled as a review. At this stage, the student can view the answer s/he gave or see the grade if the assignment is graded.

Within the scope of RQ3, the final grades of the students for the Operating Systems course were considered as an indicator of academic performance. Students took two written exams (i.e., first in the midterm and second in the final exam) during the semester. Apart from that, they received assignments regularly in Moodle during the semester. The students' final grades were calculated by taking 25% of the midterm exam, 25% of their assignment scores in Moodle, and 50% of the final exam. The final score was used in the data analysis by categorizing it as "Passed" and "Failed." The grades were categorized as "Failed" (n = 30, final score < 50) and "Passed" (n = 39, final score ≥ 50) considering the indicators in the undergraduate regulations of the university.

3.2. Data pre-processing and feature extraction

A total of 9633 activities of 69 students who submitted their assignments before the deadline are exported from Moodle's database. The log sequence for a student can include all the events given in Table 1. Also, *Assignment viewed*, *Question viewed*, and *Question reviewed* events can take place more than once in a log sequence. Among the examined records, the shortest log sequence contains only 4 records, while the longest log sequence consists of 268 records. While an average log consists of 45 records, the median value is 39. An example of a log sequence consisting of 14 records of a student is as follows: *Assignment viewed -> Attempt started -> Question viewed -> Question viewed -> Question viewed -> Question viewed ->* Question viewed -> *Question viewed -> Assignment submitted –> Assignment viewed -> Question reviewed-> Question reviewed-> Question reviewed-> Question reviewed*. During the data pre-processing, Moodle log records are processed and features were extracted for each student. This operation was repeated four times for each assignment. Description of the extracted features are given in Table 2. These features were selected in the light of existing literature (Akçapınar & Kokoç, 2020; Cerezo et al., 2017; Stiller & Bachmaier, 2019). For example; time-related features (e.g., Duration, Time taken) were selected since previous studies showed that time spent on a task is an important feature while identifying at-risk students as well as understanding their motivation and competencies in metacognitive learning strategies (Stiller & Bachmaier, 2019). Features related to procrastination behavior (e.g., Started on, Completed) were also found to be effective while clustering students based on their assignment submission behaviors (Akçapınar & Kokoç, 2020) and predicting their academic achievements (Cerezo et al., 2017).

Table 2. Features used in the study and their descriptions

Feature	Description
Attempt count	The number of time student view the questions.
Duration	The amount of time a student spends on an assignment (in minutes).
Started on	The difference between the date and time the assignment was started and the due date (in hours).
Completed	The difference between the date and time the assignment was submitted and the due date (in hours).
Time taken	The amount of time it took the student to start and submit the assignment (in hours).

3.3. Data analysis

The study used cluster analysis to group the students according to similar assignment submission behaviors. As a temporal analysis, Markov Chains were conducted to model transition behaviors of online assignment submission, and association rule mining was used to build predictive rules based on the students' behaviors and academic performances. Since the contents, question types, and the numbers of the questions are varied in different assignments, the students' assignment submission behaviors are clustered independently for each assignment. To map the clusters in different assignments, each assignment should have the same number of clusters and features. The clustering process was carried out with categorical data. Hence, all features were categorized into three levels using the equal interval method. Data analysis and visualizations were performed using the R data mining tool (R Core Team, 2017). Specifically, cluster analysis was carried out using the K-Modes algorithm with the help of the R package named *klaR*. Markov Chains analysis was performed using the *Markov Chain* package and the association rule mining analysis was performed using the *arules* package.

4. Results

4.1. What are the students' behavioral patterns of online assignment submission? (RQ1)

Within the scope of the second research question, it was investigated whether the students' homework submission behavior changed over time. For this purpose, students were divided into three clusters for each assignment independently. The number of clusters determined to be three due to the high interpretability of having *high*, *medium*, and *low* engaged clusters. Whether the three clusters solution fits the data is validated visually using the Elbow method. The scaled cluster centers' distributions formed after the cluster analysis are presented in Figure 1 for each assignment. The cluster centers showed that students displayed similar patterns in all four assignments. For example, the students in the second cluster in Assignment1 and the students in the first cluster in Assignment2, the students in the third cluster in Assignment3, and the students in the first cluster in Assignment4 displayed the same pattern. The prominent features of these students are- they start the assignment at the last moment (*StartedOn*), spent less time to complete the assignment (*Duration*), and the number of questions displayed (*AttemptCount*) is less. In other words, the students in these clusters submitted the assignment, but they gave a minimum effort for the assignment. Similarly, the students in Cluster3 in Assignment1, the students in Cluster2 in Assignment2, the students in Cluster1 in Assignment3, and the students in Cluster2 in Assignment4 also displayed a similar behavioral pattern. The prominent features of these students are- they started the assignment much earlier than the given deadline (*StartedOn*), spent more time to complete the assignment (*Duration*), there is a significant difference between the start and end time of the assignment (*TimeTaken*), and the number of question views (*AttemptCount*) is much higher than the other students. Although most of the students in these clusters complete their assignment submission on the last day, they start working on the assignment much earlier than the other students and they make much more effort to complete the assignment. Finally, it is observed that the students in Cluster1 in Assignment1, in Cluster3 in Assignment2, in Cluster1 in Assignment3, and in Cluster3 in Assignment4, exhibit similar assignment submission patterns. Like the students in the first group, these students start their assignment submission near the deadline (*StartedOn*), but they spend more time completing the assignment than the first group.

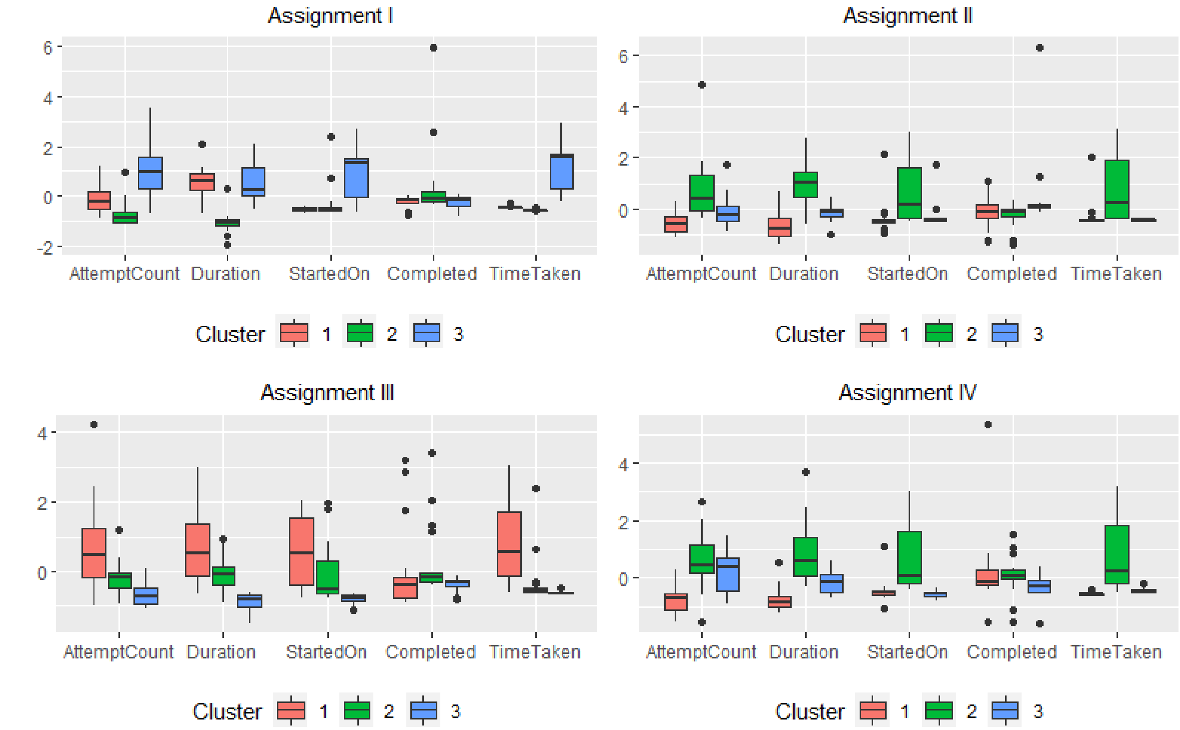


Figure 1. Box plots of features in different clusters for each assignment

In further analysis, similar clusters in each assignment were labeled as *High*, *Medium*, and *Low* in order to analyze students who followed a similar assignment submission pattern. Students who did not submit their assignments are labeled as *None*. Regarding this analysis, Cluster3 in Assignment I, Cluster2 in Assignment II, Cluster1 in Assignment III, and Cluster2 in Assignment IV are mapped to the High group. Cluster1 in Assignment I, Cluster3 in Assignment II, Cluster2 in Assignment III, and Cluster3 in Assignment IV are mapped to the Medium group. Cluster2 in Assignment I, Cluster1 in Assignment II, Cluster3 in Assignment III, and Cluster1 in Assignment IV are mapped to the Low group. Students who did not submit their assignments were manually assigned to the None group. The distribution of students in each group for all assignments are shown in Table 3.

Table 3. The number of students in each cluster after mapping

Cluster	Assignment I	Assignment II	Assignment III	Assignment IV
High	14	18	18	19
Medium	18	14	19	11
Low	16	25	13	18
None	21	12	19	21
Total	69	69	69	69

4.2. How do students' behavioral patterns of online assignment submission change over time? (RQ2)

Within the scope of the second research problem, it was investigated whether the homework submission behavior of the students changed over time. For this purpose, firstly, the transition between the sets in which the students took part in different assignments is visualized in Figure 2. As seen in the graph, there are transitions between High-Medium, Medium-High, Medium-Low, Low-Medium, Low-None, and None-Low states. On the other hand, it is also noticed that there are limited transitions between High-Low, Low-High, High-None, None-High, Medium-None, and None-Medium states. Markov Chains analysis was used to analyze the transitions between different states in more detail. In this way, the student's probabilities of transition from None, Low, Medium, or High status in one assignment to None, Low, Medium, or High status in another assignment were calculated. The values calculated for Assignment1-Assignment2, Assignment2-Assignment3, and Assignment3-Assignment4 transitions are presented in Figure 3.

As stated earlier, we clustered students in High, Medium, Low, and None after mapping their assignment submission behavior. Hence, the Markov Chain analysis in Figure 3 shows the actual transition probabilities between the groups across the semester.

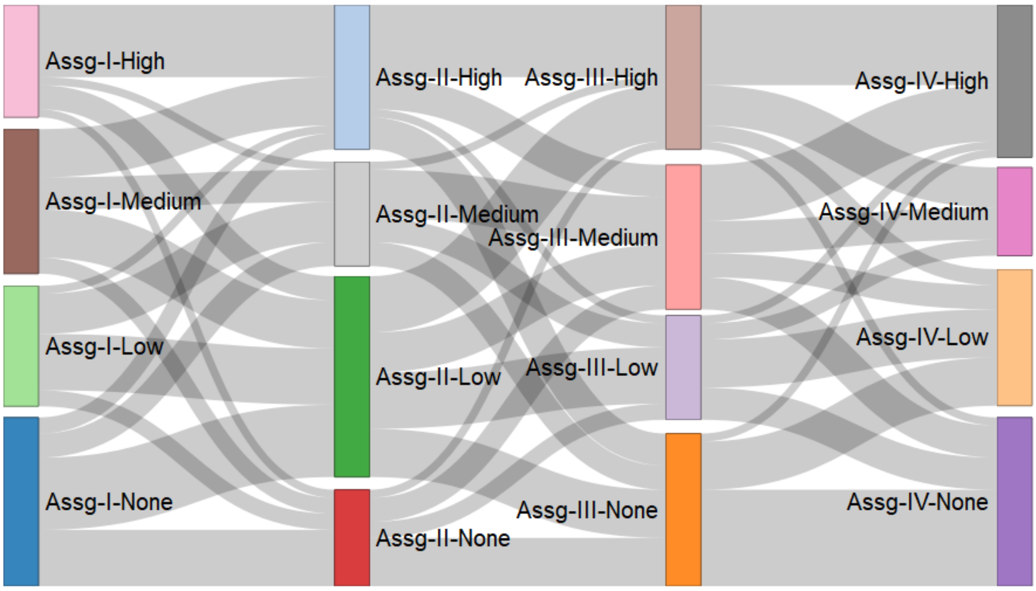


Figure 2. The students' assignment submission behaviors over time

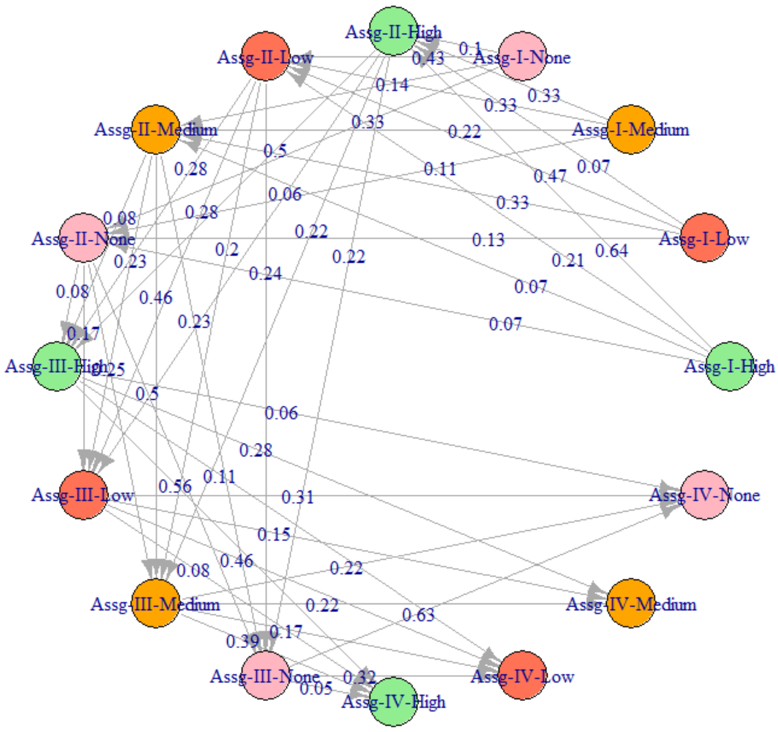


Figure 3. Transition probabilities among different assignments

The arrow between the groups indicates the direction of the transition and the numerical values represent the probability of the transition between each group. The highest probability of each transition is 1 (that is, 100%). Our Markov Chains analysis uncovered some important assignment submission behaviors of the students; therefore, we elaborate four key transitions, namely High-to-None, High-to-Low, None-to-High, and Low-to-High. For High-to-None transition, the Markov Chains analysis indicates that- students in the High cluster who submitted Assignment I have the transition probability of 0.07 to be in the None cluster in their Assignment II submission. This means the High-to-None cluster transition is like this that only 7 out of 100 students will not submit their Assignment II who belonged to the High cluster in their Assignment I submission. Consequently, for High-to-Low transition, students in the High cluster who submitted Assignment I will have a 0.21 (i.e., 21 students out of 100) transition probability to be in the Low cluster in their Assignment II submission. Similarly, for the None-to-High transition, the probability is 0.1. This means, only 10 out of 100 students who belonged to the None cluster in their Assignment I submission will be in the High cluster in their Assignment II submission. In the case of Low-to-High transition behavior, we found that the transition probability of assignment submission is 0.07 (7 out of 100 students) between Assignment I's Low cluster and Assignment II's High cluster.

4.3. What are the association rules between students' online assignment submission behaviors and their learning performance that can be used to predict at-risk students as early as possible? (RQ3)

RQ3 was answered using Association Rule Mining (ARM) analysis. The rules related to passing and failing the course were filtered among the found rules. As a result, 20 rules for students who passed the course and 14 rules for students who failed the course were obtained. The list of rules obtained and Support, Confidence, and Lift values for each rule are presented in Table 4.

Table 4. The list of the association rules extracted

No	LHS		RHS	Support	Confidence	Lift
1	{Assg-IV-Medium}	=>	{Passed}	0.16	1.00	1.77
2	{Assg-II-High,Assg-IV-High}	=>	{Passed}	0.16	1.00	1.77
3	{Assg-III-High,Assg-IV-High}	=>	{Passed}	0.14	1.00	1.77
4	{Assg-I-High,Assg-II-High}	=>	{Passed}	0.13	1.00	1.77
5	{Assg-II-High,Assg-III-High}	=>	{Passed}	0.13	1.00	1.77
6	{Assg-I-High,Assg-III-High}	=>	{Passed}	0.12	1.00	1.77
7	{Assg-I-Medium,Assg-III-High}	=>	{Passed}	0.12	1.00	1.77
8	{Assg-I-Medium,Assg-IV-High}	=>	{Passed}	0.10	1.00	1.77
9	{Assg-I-High,Assg-II-High,Assg-IV-High}	=>	{Passed}	0.10	1.00	1.77
10	{Assg-II-High,Assg-III-High,Assg-IV-High}	=>	{Passed}	0.10	1.00	1.77
11	{Assg-IV-High}	=>	{Passed}	0.26	0.95	1.68
12	{Assg-III-High}	=>	{Passed}	0.25	0.94	1.67
13	{Assg-II-High}	=>	{Passed}	0.25	0.94	1.67
14	{Assg-I-High}	=>	{Passed}	0.19	0.93	1.64
15	{Assg-I-High,Assg-IV-High}	=>	{Passed}	0.14	0.91	1.61
16	{Assg-I-Medium}	=>	{Passed}	0.22	0.83	1.47
17	{Assg-II-Medium}	=>	{Passed}	0.13	0.64	1.14
18	{Assg-III-Medium}	=>	{Passed}	0.16	0.58	1.02
19	{Assg-I-Low}	=>	{Passed}	0.13	0.56	1.00
20	{Assg-III-Low}	=>	{Passed}	0.10	0.54	0.95
21	{Assg-II-None,Assg-IV-None}	=>	{Failed}	0.13	1.00	2.30
22	{Assg-I-None,Assg-III-None,Assg-IV-None}	=>	{Failed}	0.13	1.00	2.30
23	{Assg-I-None,Assg-II-None}	=>	{Failed}	0.10	1.00	2.30
24	{Assg-II-Low,Assg-IV-None}	=>	{Failed}	0.10	1.00	2.30
25	{Assg-I-None,Assg-IV-None}	=>	{Failed}	0.20	0.93	2.15
26	{Assg-I-None,Assg-III-None}	=>	{Failed}	0.17	0.92	2.12
27	{Assg-II-None}	=>	{Failed}	0.16	0.92	2.11
28	{Assg-III-None,Assg-IV-None}	=>	{Failed}	0.16	0.92	2.11
29	{Assg-IV-None}	=>	{Failed}	0.28	0.90	2.08
30	{Assg-I-None}	=>	{Failed}	0.28	0.90	2.08
31	{Assg-I-None,Assg-II-Low}	=>	{Failed}	0.12	0.89	2.04
32	{Assg-III-None}	=>	{Failed}	0.22	0.79	1.82
33	{Assg-IV-Low}	=>	{Failed}	0.14	0.56	1.28
34	{Assg-II-Low}	=>	{Failed}	0.19	0.52	1.20

Rule 1 can be interpreted as- students belonging to the Medium cluster who submitted Assignment 4 on time will pass at the end of the semester. The confidence of this rule is found to be high (Confidence = 1.0, Support = 0.16, Lift = 1.77). Rule 2 also has a high confidence rate as equal (Confidence = 1.0, Support = 0.16, Lift = 1.77) as Rule 1, where it is established that- students in the High cluster who submitted both Assignment 2 and Assignment 4 are likely to pass at the end of the semester. Rules 3 to 10 generated by the association rule mining analysis have the same confidence (Confidence = 1.0) and lift (Lift = 1.77); however, the support values vary.

Rules 11 to 20 that represent the rules for the students who passed at the term-end differ much concerning each rule's confidence, support, and lift.

Rule 21 to 34 are for those students who are likely to fail at the end of the semester. For instance, Rule 21 suggests that- the students in the None cluster who had not submitted Assignment 2 and Assignment 4 are likely to fail in this course. Here, the confidence of our analysis is high (Confidence = 1.0) which means all the students who are following this pattern failed the course. The rule 22, 23, and 24 for the failed students have equal confidence (Confidence = 1.0) as Rule 21 revealed that- those students had not submitted Assignment 1-2 & 4, Assignment 1 & 2, and Assignment 2 & 4, respectively.

5. Discussion, conclusion, and limitations

Precision education aims to use artificial intelligence, LA, data analytics, text analytics, image analytics, and machine learning methods to solve complex educational problems that are yet to uncover in higher education. Along with LA, precision education also improves teaching quality and learning performance by identifying inattentive students in the classroom, at-risk students, potential drop-outs, and predicting final scores. By doing this, precision education aims to assist teachers in re-designing pedagogy, provide special care to those students in need, and provide timely feedback. A student's assignment submission is a complex aspect that has always been crucial for teachers to understand in order to provide timely feedback. In recent days, students are asked to submit their assignments using online platforms such as Moodle, Blackboard, and Google classroom. Teachers often find it difficult to understand how well a given assignment is prepared and submitted while using an online platform. In addition, it is difficult for the teachers to understand a student's learning process and assess the learning outcome just by looking at the logs. Therefore, we need to analyze these logs using precision education guidelines to reveal more insightful learning patterns such as how a student's online assignment submission behavior changes as the semester progress or find the association between students' online assignment submission behaviors and their final score. Finding these insightful learning patterns are important for teachers to provide quality education. To date, studies in precision education primarily emphasized online learning behaviors such as quiz-taking, content navigation, e-book reading, and video viewing. Therefore, most of the predictive models in the precision education literature are about identifying at-risk and drop-out using online interaction data such as reading behavior, content viewing behavior, slide navigation behavior, and related. However, a few studies have been found that analyzed online assignment submission behavior. In addition, to analyze the online submission behavior, most of the studies have overlooked the temporality (that is, the temporal analysis of learning interaction data). As mentioned earlier, temporal LA in precision education can bring new insightful information from online assignment submission behavioral patterns.

This study is conducted to tackle the abovementioned aspects of precision education. In this study, at first, we employed cluster analysis to profile students based on their online assignment submission behaviors; after that, we performed the Markov Chains analysis to investigate whether their patterns of online assignment submission behaviors change over time; and lastly, we applied the association rule mining method to examine the relationship between students' online submission behaviors and their course success. Although numerous studies use educational data mining methods such as clustering, regression, and classification to diagnose students' online assignment submission behaviors (Yang et al., 2020), temporal analysis has been rarely employed in educational research (Olsen et al., 2020). Thus, the study combined exploratory methods and temporal LA to extract actionable knowledge for learning designers and instructors. Our predictive models contribute to the precision education literature in terms of a deeper understanding of students' online assignment submission behavior's temporal patterns and establish the relation of these temporal patterns with their learning performances.

The first research question concerns profiling the students based on their online assignment submission behaviors. It was revealed that the students were clustered into three groups according to similar assignment submission behaviors. This result is consistent with Akçapınar and Kokoç (2020) findings, where it was found that the students' assignment submission data yielded three different clusters. Our results indicate that most of the students in cluster low and medium started their assignment submissions just before the due date. This result is likely to be related to academic procrastination behaviors. Procrastination involves delaying an assignment submission and learning task as long as possible (Yang et al., 2020). It is implied that most of the students had high procrastination tendencies based on their assignment submission behaviors. Our results are supported by previous studies indicating that time-related indicators reflected students' procrastination behaviors in online learning (Cerezo et al., 2017; Paule-Ruiz et al., 2015; You, 2016). The clusters based on the students' behaviors can be used as input to online learning environments to prevent procrastination behaviors. This predictive model can be applied to detect students' procrastination behavior from their online assignment submission behavioral data and inform the course instructor about the group of students using procrastination. Hence, our predictive model would help the instructor in planning an early intervention for those who are using procrastination regularly in an assigned learning task or a given assignment.

The second research question showed us whether the student followed the same pattern while submitting their assignments throughout the term. As a result, we found that the probability of shifting between the High and Low groups was less than 10%. We yield the conclusion that students in the High group have a low probability of going to the Low or None group. Likewise, students in the Low or None group during the beginning of the semester have a relatively low probability of going to the high group as the semester progresses. As a result, students in the Low and None group are at-risk of failing the course at the end of the semester. Nonetheless, these results support the idea that using temporal analytics provides exciting possibilities to move towards a new paradigm of assessment that replaces current point-in-time evaluations of learning states (Molenaar & Wise, 2016).

The third research question examined the relationship between students' assignment submission behavior and academic performance. The relationship was modeled using association rule mining. A total of 34 rules were generated which are related to academic performance. In practice, these rules can be used by the instructors or system designers to understand students' assignment submission patterns while the semester is in progress and to plan necessary interventions to prevent possible academic failures. Regarding the early prediction of students' end of year academic performance following rules can be used. For example, based on Rule 14 it can be speculated that if a student belongs to the High interaction group in the first assignment s/he will pass the course with a probability of 0.93. However, if s/he is in the Low group the probability of passing the course decreases to 0.56 (Rule 19). On the other hand, if s/he does not submit the first assignment (Assg-I-None) s/he will fail the course with a probability of 0.90 (Rule 30). Predictive models can be developed by using these rules to provide teachers with actionable insights to support their decision-making processes (Romero & Ventura, 2020). Thus, these rules can also be used to develop a rule-based intervention engine to prevent at-risk students, which is a core focus of precision education. The rules found in the study could be used as an input for student models in LA dashboards and intervention engines. Furthermore, researchers can use the rules to design automatic early interventions for increasing students' performance for precision education. Similarly, Tsai et al. (2020) concluded that the dropout prediction model in their study could provide early warnings and interventions to at-risk students for achieving precision education. It can be mentioned that identifying at-risk students is a key concern of precision education. Therefore, a LA intervention is required to help them to change their behaviors. By using these association rules that we generated to address RQ3, the instructor can spot the at-risk students by using the data from the first assignment (around the 4th week).

In conclusion, the main contribution of the study is unfolding students' online assignment submission behavior using temporal LA. Leveraging online assignment submission behavioral data, this study aims to contribute to precision education literature in various ways, namely by early detection of procrastination behavior, detection of at-risk students (students in Low and in None group in Figure 2), and generation of association rules for building a rule-based intervention for the course teacher. Obtained rules can be used to predict students' end-of-term academic performance from their assignment submission behaviors at the beginning of the semester. These predictive models are primarily for instructors, but students can also get benefited from them. By using the simple visualizations that have been generated by our predictive models, students can take control of their assignment submissions. For instance, if a student finds him/herself in a Low or None group in the first few weeks of the semester, s/he can step-up and quickly submit the assignment. Also, a student can control his/her procrastination behavior. Students can also find their peers who have similar behavior. The study opens up the space for future studies as well as the design and development of intervention tools based on temporal features of online assignment submission behaviors. Moreover, the study asks whether clustering analysis, temporal analysis, and association rule mining analysis could be used to explore specific patterns of assignment submission behavior. Our results indicate that temporal analysis can be used to detect the students' online assignment submission behavior patterns and transitions between related actions. This study also proves that combining various analytic methods including clustering, Markov Chains, and association rule mining is useful for modeling temporal patterns of online assignment submission behaviors. This is a methodological contribution of the study for further studies in precision education, which provides us with deeper insights into students' behavior.

This study has some limitations that need to be discussed. First, the small sample size has decreased the generalizability of our results. To overcome this, a large-scale study in the future may be conducted in the context of open online courses. Second, this study used a data-driven approach for temporal analysis of students' behaviors. Apart from online assignment submission behavioral features, other features such as the quality of assignments, learning achievement, gender, and device students used to complete learning tasks are not analyzed. In developing predictive models, it is important to analyze LMS data combining with multimodal data to understand the learning process and predictive studies (Olsen et al., 2020). Thus, in future behavior modeling studies, researchers may collect different data types from different time periods. Third, the present study did not compare procrastination tendencies, self-regulation skills, and cognitive differences of the students who have the same sequential behavioral patterns in the learning process. Therefore, future studies regarding the student modeling of online assignment submission behaviors in precision education would consider these variables.

References

Akçapınar, G., & Kokoç, M. (2020). Analyzing the relationship between student's assignment submission behaviors and course achievement through process mining analysis. *Turkish Journal of Computer and Mathematics Education (TURCOMAT), 11*(2), 386-401. doi:10.16949/turkbilmat.711683

Azcona, D., Hsiao, I., & Smeaton, A. F. (2019). Detecting students-at-risk in computer programming classes with learning analytics from students' digital footprints. *User Modeling and User-Adapted Interaction, 29*, 759–788. doi:10.1007/s11257-019-09234-7

Boroujeni, M. S., & Dillenbourg, P. (2019). Discovery and temporal analysis of MOOC study patterns. *Journal of Learning Analytics, 6*(1), 16-33. doi:10.18608/jla.2019.61.2

Cerezo, R., Sánchez-Santillán, M., Paule-Ruiz, M. P., & Núñez, J. C. (2016). Students' LMS interaction patterns and their relationship with achievement: A case study in higher education. *Computers & Education, 96*, 42-54. doi:10.1016/j.compedu.2016.02.006

Cerezo, R., Esteban, M., Sánchez-Santillán, M., & Núñez, J.C. (2017). Procrastinating behavior in computer-based learning environments to predict performance: A case study in Moodle. *Frontiers in Psychology, 8*, 1403. doi:10.3389/fpsyg.2017.01403

Chen, B., Knight, S., & Wise, A. (2018). Critical issues in designing and implementing temporal analytics. *Journal of Learning Analytics, 5*(1), 1-9. doi:10.18608/jla.2018.53.1

Cheng, H. N., Liu, Z., Sun, J., Liu, S., & Yang, Z. (2017). Unfolding online learning behavioral patterns and their temporal changes of college students in SPOCs. *Interactive Learning Environments, 25*(2), 176-188. doi:10.1080/10494820.2016.1276082

Cormack, S. H., Eagle, L. A., & Davies, M. S. (2020) A Large-scale test of the relationship between procrastination and performance using learning analytics. *Assessment & Evaluation in Higher Education, 45*(7), 1046-1059. doi:10.1080/02602938.2019.1705244

Juhaňák, L., Zounek, J., & Rohlíková, L. (2019). Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. *Computers in Human Behavior, 92*, 496-506. doi:10.1016/j.chb.2017.12.015

Knight, S., Wise, A. F., & Chen, B. (2017). Time for change: Why learning analytics needs temporal analysis. *Journal of Learning Analytics, 4*(3), 7-17. doi:10.18608/jla.2017.43.2

Kokoç, M., & Altun, A. (2019). Building a learning experience: What do learners' online interaction data imply? In D. G. Sampson, D. Ifenthaler, J. M. Spector, P. Isaias, & S. Sergis (Eds.), *Learning technologies for transforming teaching, learning and assessment at large scale* (pp. 55-70). New York, NY: Springer. doi:10.1007/978-3-030-15130-0_4

Lu, O., Huang, A., Huang, J., Lin, A., Ogata, H., & Yang, S. J. H. (2018). Applying learning analytics for the early prediction of students' academic performance in blended learning. *Educational Technology & Society, 21*(2), 220-232.

Mahzoon, M. J., Maher, M. L., Eltayeby, O., Dou, W., & Grace, K. (2018). A sequence data model for analyzing temporal patterns of student data. *Journal of Learning Analytics, 5*(1), 55-74. doi:10.18608/jla.2018.51.5

Matcha, W., Gašević, D., Uzir, N. A. A., Jovanović, J., & Pardo, A. (2019, March). Analytics of learning strategies: Associations with academic performance and feedback. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 461-470). New York: ACM. doi:10.1145/3303772.3303787

Molenaar, I., & Wise, A. (2016). Grand challenge problem 12: Assessing student learning through continuous collection and interpretation of temporal performance data. In J. Ederle, K. Lund, P. Tchounikine, & F. Fischer (Eds.), *Grand challenge problems in technology-enhanced learning II: MOOCs and beyond* (pp. 59–61). New York, NY: Springer. doi:10.1007/978-3-319-12562-6_13

Nguyen, Q., Huptych, M., & Rienties, B. (2018). Using temporal analytics to detect inconsistencies between learning design and student behaviours. *Journal of Learning Analytics, 5*(3), 120-135. doi:10.18608/jla.2018.53.8

Olsen, J. K., Sharma, K., Rummel, N., & Aleven, V. (2020). Temporal analysis of multimodal data to predict collaborative learning outcomes. *British Journal of Educational Technology, 51*(5), 1527-1547. doi:10.1111/bjet.12982

Papamitsiou, Z., & Economides, A. A. (2014). Temporal learning analytics for adaptive assessment. *Journal of Learning Analytics, 1*(3), 165-168. doi:10.18608/jla.2014.13.13

Pardo, A., & Dawson, S. (2016). Learning analytics: How can data be used to improve learning practice? In P. Reimann, S. Bull, M. Kickmeier-Rust, R. Vatrapu, & B. Wasson (Eds.), *Measuring and Visualizing Learning in the Information-Rich Classroom* (pp. 41-55). New York, NY: Routledge.

Paule-Ruiz, M. P., Riestra-González, M., Sánchez-Santillán, M., & Pérez-Pérez, J. R. (2015). The procrastination related indicators in e-learning platforms. *Journal of Universal Computer Science, 21*(1), 7-22. doi:10.3217/jucs-021-01-0007

R Core Team (2017). *R: A Language and environment for statistical computing* [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Available from https://cran.r-project.org/

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10*(3), e1355. doi:10.1002/widm.1355

Stiller, K., & Bachmaier, R. (2019). Using study times for identifying types of learners in a distance training for trainee teachers. *Turkish Online Journal of Distance Education, 20*(2), 21-45. doi:10.17718/tojde.557728

Tila, D., & Levy, D. (2020). Revising online assignments and the impact on student performance at a community college. *Community College Journal of Research and Practice, 44*(3), 163-180. doi:10.1080/10668926.2018.1564089

Tsai, S.-C., Chen, C.-H., Shiao, Y.-T., Ciou, J.-S., & Wu, T.-N. (2020). Precision education with statistical learning and deep learning: A Case study in Taiwan. *International Journal of Educational Technology in Higher Education, 17*(1), 12. doi:10.1186/s41239-020-00186-2

Wilson, M. S., & Ismaili, P. B. (2019). Toward maximizing the student experience and value proposition through precision education. *Business Education Innovation Journal, 11*(2), 119-124.

Yang, S. J. H. (2019). Precision education: New challenges for AI in education [conference keynote]. In *Proceedings of the 27th International Conference on Computers in Education (ICCE)* (pp. XXVII-XXVIII). Kenting, Taiwan: Asia-Pacific Society for Computers in Education (APSCE).

Yang, Y., Hooshyar, D., Pedaste, M., Wang, M., Huang, Y.-M., & Lim, H. (2020). Prediction of students' procrastination behaviour through their submission behavioural pattern in online learning. *Journal of Ambient Intelligence and Humanized Computing*. doi:10.1007/s12652-020-02041-8

You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *Internet and Higher Education, 29*, 23-30. doi:10.1016/j.iheduc.2015.11.003

Zacharis, N. Z. (2015). A multivariate approach to predicting student outcomes in web-enabled blended learning courses. *The Internet and Higher Education, 27*, 44-53. doi:10.1016/j.iheduc.2015.05.002

Lee, A. V. Y. (2021). Determining Quality and Distribution of Ideas in Online Classroom Talk using Learning Analytics and Machine Learning. *Educational Technology & Society*, *24* (1), 236–249.

Determining Quality and Distribution of Ideas in Online Classroom Talk using Learning Analytics and Machine Learning

Alwyn Vwen Yen Lee
National Institute of Education, Nanyang Technological University, Singapore // alwyn.lee@nie.edu.sg

ABSTRACT: The understanding of online classroom talk is a challenge even with current technological advancements. To determine the quality of ideas in classroom talk for individual and groups of students, a new approach such as precision education will be needed to integrate learning analytics and machine learning techniques to improve the quality of teaching and cater interventive practices for individuals based on best available evidence. This paper presents a study of 20 secondary school students engaged in asynchronous online discourse over a period of two weeks. The online discourse was recorded and classroom talk was coded before undergoing social network analysis and k-means clustering to identify three types of ideas (promising, potential, and trivial). The quality and distribution of ideas were then mapped to the different kinds of talk that were coded from the online discourse. Idea Progress Reports were designed and trialed to present collective and individual student's idea trajectories during discourse. Findings show that the majority of ideas in exploratory talk are promising to the students, while ideas in cumulative and disputational talks are less promising or trivial. Feedback on the design of the Idea Progress Reports was collected with suggestions for it to be more informative and insightful for individual student. Overall, this research has shown that classroom talk can be associated with the quality of ideas using a quantitative approach and teachers can be adequately informed about collective and individual ideas in classroom talks to provide timely interventions.

Keywords: Precision education, Machine learning, Learning analytics, Idea Identification and Analysis (I2A), Idea Progress Reports (IPR)

1. Introduction

In an era of unprecedented change and technological advancements, learning analytics has emerged as a nascent field that advance the understanding of learning processes (Siemens, 2013). Apart from using insights to provide teachers with timely but short-term interventions based on teaching and learning experiences, the true test in the long term would be to demonstrate how analytics can impact student learning and teaching practices (Gašević, Dawson, & Siemens, 2015). The larger and more effective goal will be to achieve personalized learning in current forms of mass public education systems while being cost-effective, which means avoiding running into Bloom's 2 Sigma problem (1984) that looked for methods of group instruction that can be as effective as personalized tutoring. Personalization of learning therefore remains a non-trivial task and although it has become more feasible with advanced technologies, efforts to maintain and scale best efforts past individual case studies of classes or schools, however, remain arduous.

The concept of precision education, as Hart (2016) explained, seeks to provide researchers and practitioners with tools to better understand complex mechanisms that hinder personalization at scale, allowing for a more effective approach to education. Similar and inspired by the Precision Medicine Initiative (Collins & Varmus, 2015), the creation of data would be necessary for gaining better understanding at the individual level, but such data is already prevalently abundant in the educational context and presents the next challenge: How to analyze and interpret an immeasurably large amount of student-related data to benefit students at the individual and micro level.

This challenge is familiar in communal discourse settings, where individual student interact, discuss, and share their ideas with each other in online discourse, creating an immense amount of textual data which traditional analytical methods have made attempts to process, albeit with partial success and trade-offs at scale. Teachers may try reading most, if not all, of the classroom discourse to gain a rudimentary understanding of student understanding but keeping track of ideas contributed by various students across the whole discourse is no mean feat. The fact that researchers have to select certain models and techniques to deal with subsets of data indicates that there is still difficulty in handling big discourse data. As part of the ongoing Fourth Industrial Revolution that has fundamentally transformed the scale, scope, and complexity of how people live, work, and study, the response to it must be integrated and comprehensive, to include all stakeholders from civil society to academia (Schwab, 2016). This industrial revolution is disruptive in almost every industry and country, led by emergent technologies such as Artificial Intelligence (AI) and other technological advancements such as learning analytics and machine learning techniques. With new affordances from novel methodologies and developments, it is now possible to review classroom data and analyze it with a contrasting perspective and under a different scope. Attention and emphasis can also be shifted from communities to individual for garnering deeper insights of how teaching practices and student learning can be improved on the individual, classroom-wide, and institutional levels.

This study adopts an "Idea Identification and Analysis" (I2A) methodology proposed by Lee, Tan, and Chee (2016) that was later improved in further iterations (Lee & Tan, 2017a; Lee & Tan, 2017b; Lee & Tan, 2017c). The I2A methodology identifies components of abstract entities such as ideas in discourse from online classroom talks, using a combination of learning analytics, social network analyses, and machine learning techniques. The resulting classification of ideas from discourse allows promising ideas in discourse to be differentiated from less promising or trivial ideas, so that teachers are able to focus on critical ideas that can advance lesson objectives in time-constrained lessons. In essence, although teachers may be conscious of different kinds of classroom talks (Mercer, 2008), they are however unable to delve deep into the discourse to gather insights of students' ideas with limited resources. The I2A methodology can be used to inform teachers about students' ideas at any point in time during an online discourse and through this study, this information can also be made available to individual student through summaries, such as an Idea Progress Report.

The research question guiding this study is: "How can learning analytics, machine learning, and Idea Progress Reports be used for determining the quality and distribution of ideas in different classroom talks to inform personalized interventions?"

2. Context and approach

2.1. Precision education as a new challenge for AI in education

Precision education is currently considered a new challenge of applying emergent technologies, such as AI, machine learning, and learning analytics for improving teaching quality and students' learning outcomes (Yang, 2019). The goals are aplenty in literature with a major focus on identifying at-risk students to provide timely interventions (e.g., Lu et al., 2018) and to enhance student outcomes through greater predictive accuracy (Kuch, Kearnes, & Gulson, 2020). The eventual objective is to tailor preventive and interventive practices to individuals based on best available evidence (Cook, Kilgus, & Burns, 2018).

Precision education in other research fields such as healthcare has moved emphasis from population-wide usage towards personalized medical care with the use of AI, such as in the field of radiology (Duong et al., 2019). Precision education has also emerged as an important aspect in the fields of policy sociology that takes into account data based on psychology, neuroscience and genomics (Williamson, 2019), as part of advocation by international organizations, such as OECD, to transmit scientific evidence into education policy and practice (Kuhl, Lim, Guerriero, & Van Damme, 2019)

In the field of education research, several studies have used context personalization (e.g., Bernacki & Walkington, 2018), by incorporating students' individual out-of-class interests into learning tasks so as to positively affect students' situational interests and their learning in mathematics. Other studies (e.g., Lin et al., 2017) have also extended this research to the field of computational thinking, by examining how customization of tools (e.g., character customization) can influence factors related effects, such as transfer, self-efficacy, and motivation. These studies were conducted as part of the hypothesis that customization can lead to higher and better learning outcome and could also provide greater flexibility for students who are less adaptable to new learning styles, thus reducing the chances of them being left behind.

2.2. The focus on ideas in discourse

Precision has been argued to require new data production and aggregation frameworks to measure and intervene, while drawing on established subjectivities to present newer insights (Kuch et al., 2020). Apart from previous approaches of developing customizable software features and handling of personal data from newer sources, it is feasible to start looking at the development of tools and methods that capture data, analyze, and present insights related to abstract entities such as ideas, which was previously not possible without state-of-the-art methodologies and techniques.

In an educational setting, individuals interact and share ideas to collaborate and build their understanding of the world, by treating ideas as real things, as objects of inquiry and improvement in their own right (Scardamalia & Bereiter, 2003). An idea, is hence, more than just a unit of thought, but rather the provision of epistemic function to represent something else with an ability to improve and extend beyond itself. When situated in a discourse, ideas can represent something pictured in mind, part of a concept, or as a way of explaining phenomenon. At the initial stage, ideas are, however, represented in preliminary forms with uncertain prospects (Chen, Scardamalia, & Bereiter, 2015). In order to achieve a higher level of understanding, ideas that are improvable and capable of moving the community in a forward direction are highly desirable and these ideas with *promisingness* (Chen, 2014) are critical for ensuring productive and effective classroom talk, especially when posed with authentic problems.

Ideas were differentiated in Lee's et al. (2016) work using three factors, namely, (a) the relevancy to the community; (b) the sustainable level of interest to the community; (c) the likely impact of the idea on discourse. The same research also defined different types of ideas in discourse, noting that promising ideas are of great relevancy to the community, sustains interests of the community, and are therefore worth pursuing. Potential ideas are relevant to some extent but suffers from waning communal interest over time, therefore requiring scaffolds and interventions to maintain communal interest. Last, trivial ideas are of minimal relevance and interest to the community.

2.3. Relating ideas to classroom talks

Mercer (2008) described in his work about talk as one potential influence on the development of students' knowledge and understanding. Talk can be used as a tool for learning and the focus should be on the quality of classroom talk, arguing that the social interactions and collaborative activities in the class can provide valuable opportunities for learning. For example, on the one hand, *exploratory talk* is defined to be a "joint, coordinated form of co-reasoning, in which speakers share relevant knowledge, challenge ideas" (Mercer, 2008, pp. 95). On the other hand, *disputational talk* consists of cycles of assertion and counter-assertion, forming sequences of short utterances that rarely include explicit reasoning (Mercer, 1995). *Cumulative talk* is the middle ground of exploratory and disputational talks, where students share some knowledge and ideas but in an uncritical manner with little evaluation.

In this study, the aim is to show that learning analytics and machine learning techniques can aid the investigation of ideas containing different levels of *promisingness* in discourse, a challenging process considering the nature of unstructured textual data. The distribution and quality of ideas can then be mapped to different kinds of talk that emerge from online classroom discourse.

2.4. Social network analysis and machine learning techniques in discourse analysis

Social network analysis (SNA) is an appropriate practice for analyzing social patterns of leaners and community structures (Scott, 1988). However, Oshima's et al. (2007) work found that SNA may be insufficient for examining community knowledge advancement through students' collaboration and interaction networks. In order to focus on the patterns of emerging ideas in discourse, the I2A methodology involves the generation of social networks based on bipartite relationships that associate keywords, discourse participants, and discourse units, which are then used to calculate the network measures of the discourse unit network. The study of indicators such as "centrality" determine the level of interaction between students (Wortham, 1999). Among common methods of measuring centrality, this study uses two conventional network measures, namely the betweenness centrality (BC) and the degree centrality (DC). The role of BC for any given node refers to the degree of importance of the node in helping to connect ideas in a discourse, whereas the DC is a good measure of connectivity from the node to the rest of the network.

Both BC and DC measures were similarly used in a separate study (Oshima, Oshima, & Fujita, 2016) to distinguish epistemic actions for awareness of lack of knowledge in students. For this study, the goal is to aid the identification of promising ideas from classroom discourse. The process whereby participants share and exchange information often leads to the creation of meaningful links between normal communicative speech and usage of important keywords. Therefore, since ideas are considered to be central to discussions and for mediating opinions among students, the examination of the BC and DC measures can provide insights on the degree of sharing and level of communication by students within a discourse network.

In recent years, learning analytics and machine learning were more frequently used for analyzing discourse. Examples include Discourse-Centric Learning Analytics (DCLA; Knight & Littleton, 2015) and methods that are either semantic based (e.g., Hsiao & Awasthi, 2015) or involve topic models, such as those built on Latent Dirichlet Allocation (LDA; Blei, Ng, & Jordan, 2003). Extended variations include structural topic modelling, which integrates computer-assisted text processing (Roberts, Stewart, & Tingley, 2013). Alternative methodologies have also emerged in recent times and are able to process multi-dimensional data. Examples of these methodologies include machine learning techniques for automatic text classification (e.g., Garrard, Rentoumi, Gesierich, Miller, & Gorno-Tempini, 2014), clustering techniques with Part-of-Speech (POS) tagging (e.g., Owoputi et al., 2013; Lamar, Maron, Johnson, & Bienenstock, 2010), and Natural Language Processing (NLP) related methods. With multiple sources of data and features to choose from, these studies have narrowed their focus to specific types of context, data, and instruments, in order to make sense of the data and analyse the various impacts on learning. Only a few methodical approaches, including the approach taken in this study, attempt to conduct idea analysis and discern the quality of ideas in discourse to further understand how classroom talk can be associated with the quality of ideas to adequately inform teachers and provide timely interventions.

2.5. Idea Progress Report as teacher feedback to individual student

Prior research (e.g., Tunstall & Gsipps, 1996; Van den Bergh, Ros, & Beijaard, 2012) have shown that teacher feedback to students is crucial for enhancing and progress in student learning. Teacher feedback can be verbal or written and exists in different forms, such as reports, rubrics (Wollenschläger et al., 2016), or corrective responses (Zheng & Yu, 2018). Teacher feedback should be related to meta-cognition, social learning, and learning goals but such instances are still rather scarce in practice (Van den Bergh et al., 2012). In this study, to ensure that teachers are provided with sufficient information to cue critical and timely interventions for students, an artifact in the form of an Idea Progress Report (IPR) was designed and trialed.

The IPR is a feedback tool that was designed with precision education in mind. The goal was to provide adequate information and assistance to the teacher to make informed and timely interventions for students who were participating in knowledge building activities within the classroom. In preparation for this study, data from prior research was reorganized and analyzed based on the I2A methodology (Lee et al., 2016; Lee & Tan, 2017a) and with consideration of Mercer's three kinds of talks (2008).

3. Methodology

3.1. Participants

A total of 20 secondary school (Grade 8) students were involved in this study under the instruction of an experienced teacher. He was pedagogically trained and has been teaching for nearly a decade at the point of the study and was able to facilitate computer-aided and knowledge building lessons effectively.

3.2. Dataset and settings

The dataset in this study includes 101 notes written by students on the Knowledge Forum (Scardamalia, 2004), an online discourse platform that supports knowledge building. These notes are online postings written by the students, consisting of their ideas, discussions, and arguments about an authentic problem, which is "how and why is an uncle suffering from cardiovascular problems" and related to the science topic on the "human circulatory system." The focus of this study is the textual content of the notes, which was extracted, cleaned, and anonymized to protect the identities of students. The online discourse was recorded over a period of two weeks and held at a computer-aided environment in the same secondary school.

3.3. Determining discourse groups and types of talk in online discourse

To determine the kind of talks that could be present in an online and asynchronous discourse, a virtual space on the Knowledge Forum was hosted for students to contribute and build on each other's responses (see Figure 1).

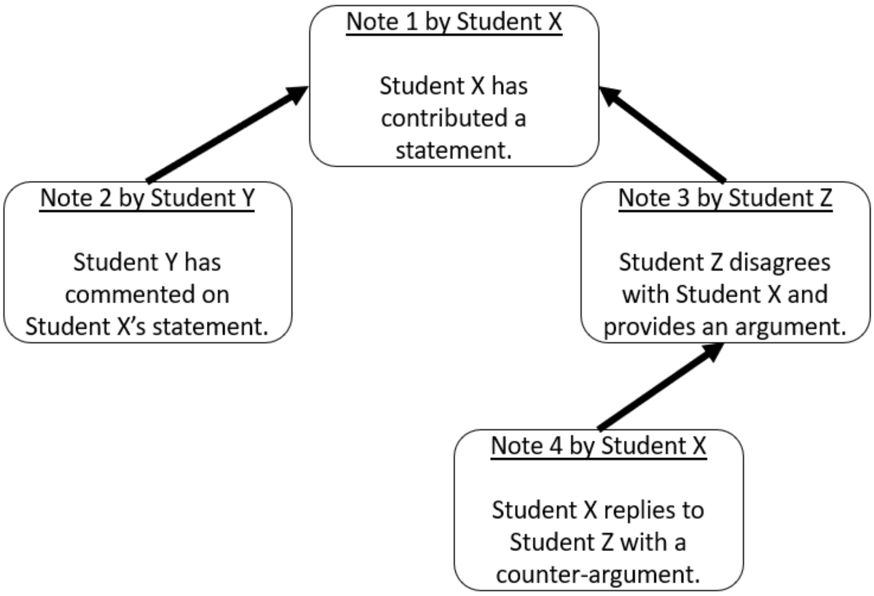


Figure 1. Example of how students build on each other's statements, claims, or ideas in Knowledge Forum

The replying and quoting mechanism of threaded discussions is commonly seen on discussion boards and forums, often presented in a top-down and linear format. When students are provided with a virtual space (also known as a "view") on the Knowledge Forum, students could read other student-written notes on the same view and visually estimate the width and depth of discussion as they participate, without clicking into the discussions. Since Knowledge Forum notes are movable features on the virtual space, students could move their notes around the virtual space to form separate discussion groups with their own peers. This feature does not impact the overall quality of classroom discourse and further enabled analysts to visualize and spatially identify discourse groups in classroom discourse, different from conventional methods that may need to pre-define discourse groups through assignment or conversational analysis of turn-taking.

To illustrate this point, Figure 2 shows a screenshot of how discourse groups (circled and labelled) can be visually identified on a Knowledge Forum view, based on how threads are being initiated and continued with notes building on each other (represented by single-headed arrows) by multiple students. The talks in these individual discourse groups were then examined and qualitatively coded based on Mercer's classification of talk (Table 1).

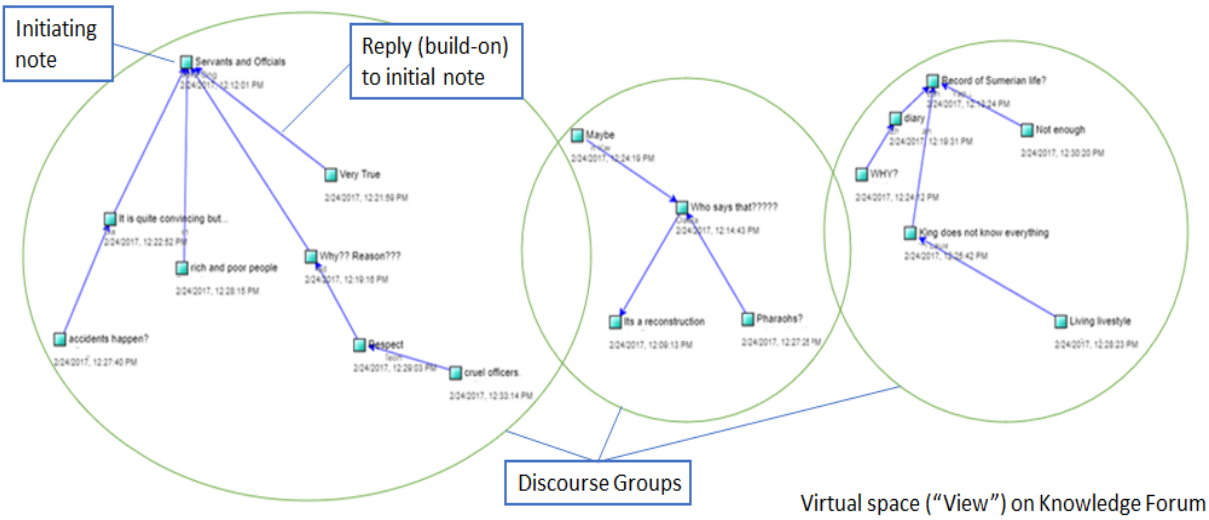


Figure 2. A screenshot on Knowledge Forum showing how students can estimate the depth of discussion at a glance and how analysts can visually identify discourse groups

Table 1. Mercer's (2008) classification of classroom talks and examples of sequences within the talks

Types of classroom talk	Observations	Examples of sequences from this study's dataset (Students' names are represented by alphabets)
Disputational	• Cycles of assertion and counter-assertion, forming sequences of short utterances • Little effort to pool resources or offer constructive ideas • Competitive instead of cooperative environment	A: "What happens if it is normal?" B: "What is it?" A: "Like, it is a verb, that's all." B: "What does it refer to?" A: "Ok, fine! It refers to the pain."
Cumulative	• Students are accepting of other ideas but in an uncritical manner • Some sharing of knowledge is present with some build on • Repetition of each other's ideas with little evaluation involved	A: "What will be the consequences if he continues to eat unhealthily?" B: "His arteries will keep on collecting fats." C: [quotes B's reply] "Uncle's arteries will continue on collecting fats and his arteries will become blocked and it will burst." D: "My theory – His condition will definitely get worse if he continues eating unhealthily."
Exploratory	• Students actively listen to each other and share ideas • Ideas may be challenged with reasons • Joint, coordinated form of co-reasoning	A: "How will the operation be done?" B: "A balloon will be inserted into his coronary artery and inflated…" C: "All these procedures increase blood supply to your heart but they do not cure coronary heart disease" A: "So what cures coronary heart disease?" C: "Treatments include lifestyle changes, medicines, medical procedures."

3.4. Idea Identification and Analysis using social network analysis and clustering technique

The Idea Identification and Analysis (I2A; Lee et al., 2016) methodology was conducted to identify and classify ideas in the online discourse, serving as an indication of communal understanding and a quantifiable measure related to the quality of ideas that were proposed and discussed by the students. The methodology is split into two phases.

The first phase involves text mining to discover keywords that are basic units of analysis that can also indicate partial resemblance of ideas when present in groups, phrases, or sentences. A text miner (Reategui, Epstein, Lorenzatti, & Klemann, 2011), based on the work of Schenker (2003), was adapted for educational purposes and mines the textual discourse data to generate a list of related conceptual keywords. To enhance the accuracy of the miner, an in-built thesaurus was included to ensure that stop words, noun markers (e.g., determiners like "the," "this"), pronouns such as "his," "her," and non-unique synonyms (e.g., using "student" to represent "students," "pupils," and "children") are excluded from the final list of keywords. The resulting list of keywords would then serve as inputs for the second phase of the methodology.

The second phase of I2A utilizes a mixture of social network analysis and machine learning to pinpoint the location of ideas in the discourse and determine the quality of ideas via unsupervised learning. A social network analyzer (KBDeX; Oshima, Oshima, & Matsuzawa, 2012) was used to generate social networks based on bipartite relationships that associate keywords, discourse participants, and the discourse units, which in this study, refer to notes written by students on the Knowledge Forum. A discourse unit (DU) may exist as a standalone note containing statements or claims written by students but can also be found as part of a threaded sequence as shown in Figure 1. For example, a student who posted new information in a note (DU1: "Clogged arteries result from a buildup of a substance called plaque on the inner walls of the arteries…"), was built on with a following note (DU5: "I need to understand – how does this affect us?"). These DUs are often chronologically labelled according to the time of posting to the discourse space. The relationships among the social networks are then analyzed to calculate conventional network measures, which in this study refers to the betweenness centrality (BC) and degree centrality (DC). These two network measures are utilized together in this study, resulting in the reorganization of discourse data onto a two-dimensional variable space plot. This variable space plot containing the network measures is referred to as the DC-BC graph and provides a discourse unit visualization overview, where discourse units are shown side-by-side on a same plot during any point in time of the discourse. This plot is shown in the findings and provides a visual method of estimating the *promisingness* of ideas.

The k-means clustering algorithm can use the same plot to determine idea quality, by being implemented to the DC-BC graph to form "k" number of clusters that represent the three types of ideas present in this study. While the use of other similar machine learning algorithms such as the supervised k-nearest neighbor (k-NN) algorithm was contemplated, the k-means algorithm was selected due to its unsupervised nature and with consideration of the limited datasets available in this study. This study focused on three likely types of ideas in discourse and by using Euclidean distance as the distance metric in determining the global silhouette peak value at an optimal "k" value, the value of "k" in this study was determined to be three, allowing the final clustering results to show the type of ideas that likely reside within individual DUs. Last but not least, the categories of ideas estimated via clustering were qualitatively analyzed and verified.

3.5. Qualitative determination of idea quality in various online classroom talk

Based on Mercer's classification of talk (Table 1), the talks in the discourse groups were qualitatively coded by two researchers who have extensive experience in working with knowledge building discourse. Both researchers were first provided with a list of discourse groups, with each group containing a thread of student-written notes. The notes were then qualitatively scrutinized to determine if there are observations or indications in the talk similar to the observations and examples shown in Table 1, which are then used as evidence for labelling the discourse group accordingly. The resulting inter-rater reliability was calculated with the remaining differences between the two raters resolved after further discussion, culminating in a final determination of the different kinds of talk in the knowledge building discourse. After the various discourse groups were qualitatively coded and I2A was conducted on the discourse data, the quality of ideas that was determined through the k-means clustering technique was mapped to the coded classroom talk. This provided a sense of how different quality of ideas in discourse may influence or lead to certain kinds of classroom talk, and whether students will be inspired or discouraged from building on each other's ideas.

3.6. Design of Idea Progress Report

After the distribution of ideas in DUs was determined, this information was provided to the teacher as feedback in the form of an Idea Progress Report (IPR). The IPR served as a one-page summary describing statistics and details of a student's idea trajectory as a member of the discourse community. The various sections of the IPR that are shown in Figure 3 provide critical details for informing students and to recognize their efforts in the crafting and dissemination of ideas within the community. These details may include a profile of an individual student, the date-time and statistics of knowledge building efforts on the Knowledge Forum, a graphical representation showing the student's efforts relative to other students, and several pointers that highlight the important contributions to the community.

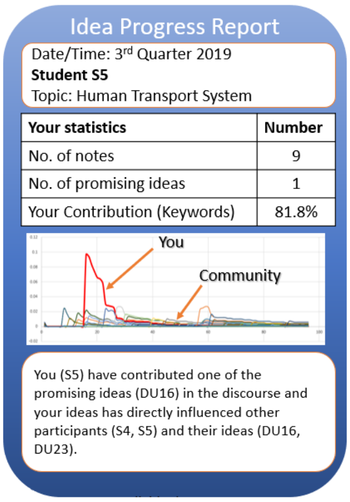


Figure 3. A screenshot of the one-page Idea Progress Report that can be issued to individual student, based on their efforts and work on ideas during the online discourse

4. Findings and discussion

This section details findings from the study in a sequential manner. First, results from the qualitative coding of the discourse groups is shown, followed by the results from the implementation of the k-means clustering on a variable space plot containing the network measures from all of the DUs in the discourse. A condensed qualitative analysis is presented for selected DUs to explain how ideas in the discourse are classified in a certain manner, before lastly, an aggregated breakdown of ideas that are found in different classroom talks is listed with suggestions on possible reasons why some types of ideas are in found in different kinds of classroom talk.

4.1. Coding of discourse groups

A total of 13 discourse groups, with the smallest group consisting of at least two notes, were identified on the Knowledge Forum view. These groups comprise 50 of the 101 notes on the Knowledge Forum view, with the remaining 51 notes belonging to standalone notes that contain claims or statements from individual student that were not built on by other students or included as part of discourse groups. The notes in the discourse groups were qualitatively coded between two expert researchers with an inter-rater reliability rate of 84.6%, whereby the remaining differences between the two raters were resolved after further discussion, culminating in a final and qualitative determination of labels for various kinds of classroom talk in knowledge building discourse.

There were altogether six instances of exploratory talk, five instances of cumulative talk, and two instances of disputational talk, as shown in Table 2. Considering that the series of lessons over the two weeks were constructed to give students opportunities to build knowledge, the overall atmosphere was conducive for sharing of ideas and the environment was purposefully constructed to be psychologically safe so that students are able to propose and share ideas freely without fear of assessments or repercussions. Therefore, the chances of encountering large amounts of exploratory and cumulative talk was not unexpected. The final section of the findings (Table 3) shows a breakdown of the distribution and quality of ideas for the different kinds of talk that surfaced in the discourse.

Table 2. Number of notes (DUs) in each discourse group and how each group was coded

Discourse group no.	1	2	3	4	5	6	7	8	9	10	11	12	13
Number of DUs in group	3	3	2	5	4	3	10	5	4	2	2	5	2
Coded talks	C	E	C	C	E	C	D	E	E	C	E	D	E

Note. The three types of talks were coded using the first letter of each type of talk, namely, "E" for Exploratory, "C" for Cumulative, and "D" for Disputational.

4.2. Classification of ideas from clustering of discourse units

Once the social network analysis was used to calculate the pairs of DC and BC values for all discourse units, the pairs of values were then plotted on the variable space plot in preparation for k-means clustering. Figure 4 shows the position of the markers on a single variable space plot, with each marker (represented by a hyphen) representing an estimate of the quality of ideas in the individual discourse units. For example, a discourse unit in the top right corner of the DC-BC graph with relatively higher DC and BC values indicates that the discourse unit is likely to contain ideas that are promising, as compared to a discourse unit at the bottom left of the DC-BC graph, which is likely to contain trivial ideas.

The k-means clustering was subsequently conducted to confirm the estimates of the discourse units. Three clusters are formed as shown in Figure 5, suggesting that the three separate clusters could be labelled as containing promising, potential, and trivial ideas respectively. From the clustering results, the blue crosses represent discourse units containing promising ideas, with the green diamonds representing discourse units containing potential ideas, and the red dashes indicating discourse units that contain trivial ideas. The centroids created due to the clustering technique also help to visualize group-centric positions of the three clusters within the entire discourse. By using the clustering technique, the quality of ideas in each discourse unit was determined in a timely and possibly scalable manner.

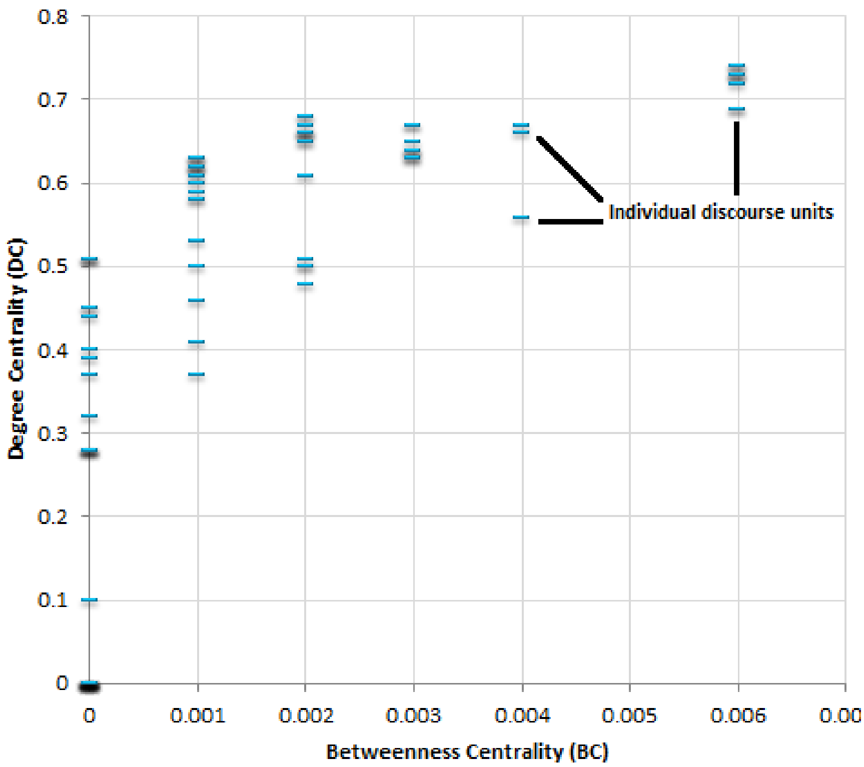


Figure 4. A DC-BC graph showing the estimated *promisingness* of ideas in each discourse unit at the end of the discourse

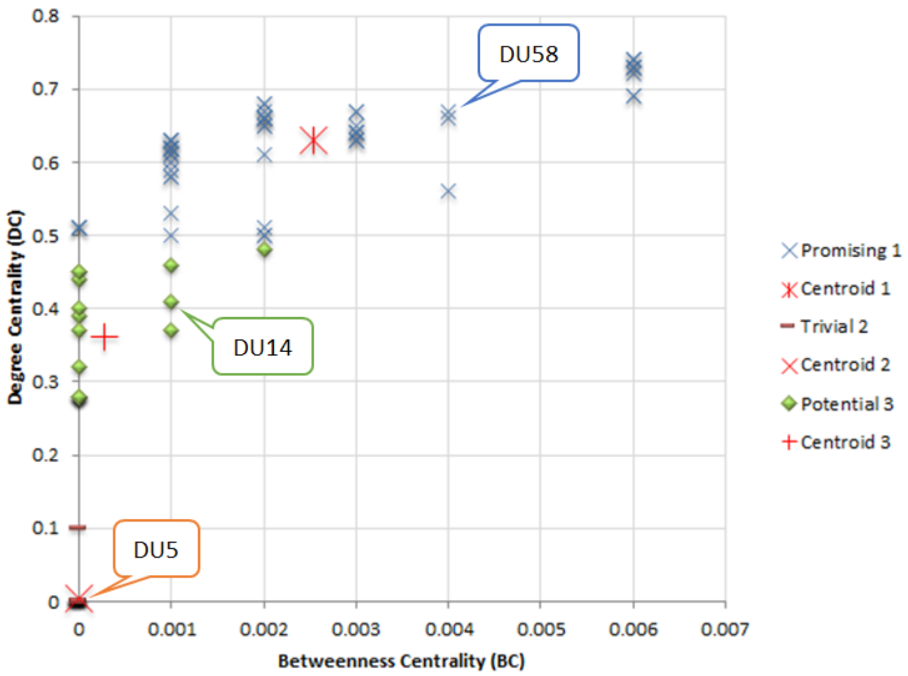


Figure 5. A DC-BG graph showing discourse units positioned in three clusters after k-means clustering was conducted on the discourse units at the end of discourse

4.3. Verification of idea quality in discourse units using qualitative analysis

Since training data was not utilized in this study due to the unsupervised nature of k-means clustering, it was noted that instead of a k-fold cross validation, a qualitative analysis was conducted to verify the quality of ideas against the qualitative content in the discourse units. The following are excerpts of the qualitative analysis for three discourse units (annotated on Figure 5) that were determined to contain trivial, potential, and promising ideas respectively.

Starting with DU5, this was a question contributed by student S3, who was trying to understand, “How does this affect us?” The reference of “this” refers to a set of new information contributed by student S1 prior to the query, which by itself was considered to contain promising ideas. However, due to the lack of relevancy and interest or impact on the discourse, DU5 was considered to be written solely as an attempt to seek clarity and was trivial.

In response to a new inquiry, another student S6 asked “Is the plaque blocking the coronary artery the same as the plaque in the teeth?” and DU14 was a response by student S1 who answered “Surprisingly, there is no link between dental plaque and the plaque build-up that attaches to your arteries. If there was, then each swallow and sip would be killing you slowly.” This discourse unit was an interesting take on how students respond to each other with some relevant information and managed to sustain some interest among the group of students for a period of the discourse, resulting in it being considered to contain potential ideas that can be further built on.

Promising ideas in discourse, such as ideas found in DU58 and contributed by student S5, sought to build on previous replies and improve on each other’s ideas. The ideas in DU58 improved on a current theory in DU56 that “cigarette smoking increases the risk of coronary heart disease by itself and that smoking increases blood pressure, decreases exercise tolerance and increases tendency for blood to clot”, by proposing a better theory that “it is the buildup of fatty material (atheroma) which narrows the artery that can cause angina, a heart attack or a stroke.” Examples of such promising ideas are often relevant to previous inquiries and context, sparking students’ interests and encourages the sustenance of discussion that impacts subsequent discourse over a longer period of time.

Overall, the presence of different types of ideas in the individual discourse units can be determined using the DC-BC graph and k-means clustering, with findings showing that identified promising ideas can be further build on to sustain knowledge building discourse.

4.4. Identifying breakdown of ideas in different online classroom talk

The findings from the clustering results were subsequently used to form a breakdown of ideas in the various types of classroom talks. Table 3 presents a breakdown of the distribution and quality of ideas that emerged from different types of online classroom talk.

Table 3. A breakdown of the distribution and quality of ideas in different types of classroom talk

Types of classroom talk	Types of ideas		
	Promising	Potential	Trivial
Disputational	6.7%	33.3%	60.0%
Cumulative	46.7%	20.0%	33.3%
Exploratory	75.0%	5.0%	20.0%

From the findings in Table 3, there is a clear split between the types of ideas that exist in different types of classroom talk. Disputational talk tends to contain the least amount of promising ideas whereas three out of four ideas in exploratory talk are considered promising. Most of the trivial ideas in the discourse also exist in disputational talk, while cumulative talk contains some of each type of ideas, with nearly half of the ideas being promising.

Looking past the numerical statistics, examples of each type of classroom talk were examined with a qualitative lens using the coded labels. For example, discourse group 12 was coded as disputational talk, consisting of mostly short interactions, statements, and agreements without explanations, such as “What is it?” “It is a verb, that’s all” and “Ok fine.” The content from the discourse units in discourse group 12 revealed little attempts by students to work together or to share their ideas, representing a dearth of promising and potential ideas.

An example of cumulative talk was found in discourse group 4, which started off with a question about the consequences of a person who continues to eat unhealthily. Students then shared their ideas and knowledge, but were mostly doing so in an uncritical manner, by simply building onto existing ideas by adding theories or opinions without evaluation. At times where comparison of theories or opinions were conducted by students in the discourse group, the evaluations then transited into part of a potential or promising idea that further encouraged discussions among the discourse group, thus reflecting the fair share of promising and potential ideas that exist in cumulative talk.

Discourse group 5 is an example of why exploratory talk tends to consist a majority of promising ideas. The talk was initiated with the proposal of a theory and relevant explanations about the functionalities of red blood in the human body. The proposal was then built on with external resources such as website links that students could use to aid their understanding. Some students decided to challenge the ideas in the first note (theory proposal) and compare with their own ideas, while other students in the discourse group deliberated and expanded on specific functionalities of blood components, leading to a better general understanding of blood components. There was a culture of respect for each other's ideas that was similarly observed in other discourse groups coded as exploratory talk, such as groups 2, 8, 9, 11, and 13. There was also no bashing of opinions, so students were more at ease to share relevant information and everyone was encouraged to contribute to the exploratory talk.

In summary, it is likely that the presence of promising ideas contributed to exploratory and cumulative talk that can be productive for students, while disputational talk contains the bulk of trivial ideas. However, the latter kind of disputational talk cannot be entirely discounted and ignored because exploratory and cumulative talk cannot be expected to occur at every turn of discourse. Therefore, as evidence in this study has also shown, promising ideas do still occur in disputational talk and teachers should take note not to totally ignore this kind of talk in the classroom but continue to monitor the discourse with discretion.

4.4. The trial of Idea Progress Report in this study

Given the breakdown of the distribution and quality of ideas for different talks, the teacher can recognize ideas and discern the level of understanding in a classroom from a collective or individual point of view. Moving one step ahead, the deployment of the Idea Progress Report (IPR) can provide a more precise and customized level of student information for teachers to provide personalized actions or interventions. The IPR prototype was trialed in this study at short notice and therefore, the IPR was not fully deployed as part of the lesson plan due to time constraints. Instead, the teacher presented the IPR to students as an optional source of information during lessons and feedback on how the prototype can be improved was garnered from both teacher and students.

The teacher was adamant that the design and maintenance of two versions of the IPR, namely the individual version for students and a collective version for the whole class, will be beneficial for teachers who do not have the time to analyze the whole discourse and for students who prefer a collective view of the entire discourse. Contrarily, some students felt that the collective version of the IPR was not useful to them but agreed that it would eventually be useful for the advancement of communal interests that may have a trickle-down effect of benefits for the students. These feedbacks are considered for improving the design of the Idea Progress Reports so as to make it more informative and insightful for both teachers and individual student in future studies.

4.5. Limitations of current study

Several limitations are acknowledged in this study. First, the provision of the breakdown, which shows the distribution and quality of ideas in various classroom talks, is considered to be indicative of possible predictive trends and not a direct representative of larger class sizes that can be easily replicated. Further, since the findings are based on a sample size of a single class discourse, the results may not be definitive at this point, but it is evidence-based and initial findings from other ongoing work has shown indications that this is an area that is worthy of exploring at scale. In line with efforts to encourage timely interventions and to provide students with personalized feedback, the work with teachers on the IPR will be continued to ensure that the reporting tool can be deployed within the limited time frame of lessons and curriculum, so that on-site data collection for idea analysis and IPR can be conducted in parallel during future studies.

5. Conclusion

In this paper, the concept of precision education has been applied to provide researchers and practitioners with tools to better understand complex mechanisms that provide a more effective approach to the understanding of ideas in online discourse, specifically during online classroom talk. Using social network analysis, learning analytics, and machine learning such as clustering techniques, three types of ideas were identified throughout discourse and were mapped to determine the distribution and breakdown of ideas in different classroom talks. The design and trial of Idea Progress Reports in the classroom also helped to highlight the fact that ideas in discourse can be used to inform teachers to deploy interventions for students who might be falling behind in lessons.

An analytical study of this nature is inconceivable a couple of decades ago, but with the emergence of nascent fields such as learning analytics and machine learning, it is inevitable that newer fields of research and methods can support novel forms of analysis and provide deeper insights on previous data that were almost impossible to process. From this empirical study, evidence have shown that it is possible to demonstrate how classroom talks can be associated with the quality of ideas in a quantitative manner. More so, this research has the potential to be used for predictive purposes in other aspects of precision education.

Acknowledgement

The views expressed in this paper are the author's and do not necessarily represent the views of the host institution. Appreciation goes to the teachers and student participants involved in this study and the use of data from the project (NRF2012-EDU001-EL008).

References

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*, 993-1022. doi:10.5555/944919.944937

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4-16. doi:10.3102/0013189X013006004

Bernacki, M. L., & Walkington, C. (2018). The Role of situational interest in personalized learning. *Journal of Educational Psychology, 110*(6), 864-881. doi:10.1037/edu0000250

Chen, B. (2014). *Promisingness judgements as facilitators of knowledge building in elementary science* (Unpublished doctoral dissertation). University of Toronto, Toronto, Canada.

Chen, B., Scardamalia, M., & Bereiter, C. (2015). Advancing knowledge-building discourse through judgments of promising ideas. *International Journal of Computer-Supported Collaborative Learning, 10*(4), 345-366. doi:10.1007/s11412-015-9225-z

Collins, F. S., & Varmus, H. (2015). A New initiative on precision medicine. *The New England Journal of Medicine, 372*(9), 793-795. doi:10.1056/NEJMp1500523

Cook, C. R., Kilgus, S. P., & Burns, M. K. (2018). Advancing the science and practice of precision education to enhance student outcomes. *Journal of School Psychology, 66*, 4-10. doi:10.1016/j.jsp.2017.11.004

Duong, M. T., Rauschecker, A. M., Rudie, J. D., Chen, P. H., Cook, T. S., Bryan, R. N., & Mohan, S. (2019). Artificial intelligence for precision education in radiology. *The British Journal of Radiology, 92*(1103), 20190389. doi:10.1259/bjr.20190389

Garrard, P., Rentoumi, V., Gesierich, B., Miller, B., & Gorno-Tempini, M. L. (2014). Machine learning approaches to diagnosis and laterality effects in semantic dementia discourse. *Cortex, 55*, 122-129. doi:10.1016/j.cortex.2013.05.008

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64-71. doi:10.1007/s11528-014-0822-x

Hart, S. A. (2016). Precision education initiative: Moving toward personalized education. *Mind, Brain, and Education, 10*(4), 209-211. doi:10.1111/mbe.12109

Hsiao, I. H., & Awasthi, P. (2015). Topic facet modeling: Semantic visual analytics for online discussion forums. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 231-235). New York, NY: ACM. doi:10.1145/2723576.2723613

Knight, S., & Littleton, K. (2015). Discourse-centric learning analytics: Mapping the terrain. *Journal of Learning Analytics, 2*(1), 185-209. doi:10.18608/jla.2015.21.9

Kuch, D., Kearnes, M., & Gulson, K. (2020). The Promise of precision: Datafication in medicine, agriculture and education. *Policy Studies, 41*(5), 527-546. doi:10.1080/01442872.2020.1724384

Kuhl, P. K., Lim, S. S., Guerriero, S., & Van Damme, D. (2019). *Developing minds in the digital age: Towards a science of learning for 21st century education.* Paris, France: OECD Publishing.

Lamar, M., Maron, Y., Johnson, M., & Bienenstock, E. (2010). SVD and clustering for unsupervised POS tagging. In *Proceedings of the ACL 2010 Conference Short Papers* (pp. 215-219). Uppsala, Sweden: Association for Computational Linguistics.

Lee, A. V. Y., Tan, S. C., & Chee, K. J. K. (2016). Idea Identification and Analysis (I2A): A search for sustainable promising ideas within knowledge-building discourse. In Looi, C. K., Polman, J. L., Cress, U., and Reimann, P. (Eds.), *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences (ICLS) 2016, Volume 1* (pp. 90-97). Singapore: International Society of the Learning Sciences. Retrieved from https://repository.isls.org/handle/1/103

Lee, A. V. Y., & Tan, S. C. (2017a). Promising ideas for collective advancement of communal knowledge using temporal analytics and cluster analysis. *Journal of Learning Analytics, 4*(3), 76-101. doi:10.18608/jla.2017.43.5

Lee, A. V. Y., & Tan, S. C. (2017b). Temporal analytics with discourse analysis: Tracing ideas and impact on communal discourse. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 120-127). New York, NY: ACM. doi:10.1145/3027385.3027386

Lee, A. V. Y., & Tan, S. C. (2017c). Understanding idea flow: Applying learning analytics in discourse. *Learning: Research and Practice, 3*(1), 12-29. doi:10.1080/23735082.2017.1283437

Lin, L., Parmar, D., Babu, S. V., Leonard, A. E., Daily, S. B., & Jörg, S. (2017). How character customization affects learning in computational thinking. In *Proceedings of the ACM Symposium on Applied Perception* (pp. 1-8). New York, NY: ACM. doi:10.1145/3119881.3119884

Lu, O. H., Huang, A. Y., Huang, J. C., Lin, A. J., Ogata, H., & Yang, S. J. (2018). Applying learning analytics for the early prediction of students' academic performance in blended learning. *Educational Technology & Society, 21*(2), 220-232.

Mercer, N. (1995). *The guided construction of knowledge: Talk amongst teachers and learners*. Cleveland, OH: Multilingual Matters.

Mercer, N. (2002). Developing dialogues. In G. Wells & G. Claxton (Eds.), *Learning for life in the 21st century: Sociocultural perspectives on the future of education* (pp. 141-153). Oxford, UK: Blackwell. doi:10.1002/9780470753545.ch11

Mercer, N. (2008). Talk and the development of reasoning and understanding. *Human development, 51*(1), 90-100. doi:10.1159/000113158

Oshima, J., Oshima, R., & Fujita, W. (2016). Refinement of semantic network analysis for epistemic agency in collaboration. In Looi, C. K., Polman, J. L., Cress, U., and Reimann, P. (Eds.), *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences (ICLS) 2016, Volume 2* (pp. 1191–1192). Singapore: International Society of the Learning Sciences.

Oshima, J., Oshima, R., & Knowledge Forum Japan Research Group. (2007, July). Complex network theory approach to the assessment on collective knowledge advancement through scientific discourse in CSCL. In Chinn, C. A., Erkens, G., & Puntambekar, S. (Eds.), *The Computer Supported Collaborative Learning (CSCL) Conference 2007* (pp. 563-565). New Brunswick, NJ: International Society of the Learning Sciences. Retrieved from https://repository.isls.org//handle/1/3406

Oshima, J., Oshima, R., & Matsuzawa, Y. (2012, June). Knowledge building discourse explorer: A Social network analysis application for knowledge building discourse. *Educational Technology Research and Development, 60*(5), 903–921. doi:10.1007/s11423-012-9265-2

Owoputi, O., O'Connor, B., Dyer, C., Gimpel, K., Schneider, N., & Smith, N. A. (2013, June). Improved part-of-speech tagging for online conversational text with word clusters. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 380-390). Atlanta, GA: Association for Computational Linguistics. Retrieved from https://www.aclweb.org/anthology/N13-1039

Reategui, E., Epstein, D., Lorenzatti, A., & Klemann, M. (2011). Sobek: A Text mining tool for educational applications. In *Proceedings of the International Conference on Data Mining* (pp. 59–64). Las Vegas, NV: The World Congress in Computer Science, Computer Engineering, & Applied Computing.

Roberts, M. E., Stewart, B. M., Tingley, D., & Airoldi, E. M. (2013, December 10). *The Structural topic model and applied social science* [PowerPoint slides]. Harvard University. Retrieved from https://mimno.infosci.cornell.edu/nips2013ws/slides/stm.pdf

Scardamalia, M. (2004). CSILE/Knowledge Forum®. In *Education and technology: An Encyclopedia* (pp. 183-192). Santa Barbara, CA: ABC-CLIO.

Scardamalia, M., & Bereiter, C. (2003). Knowledge building. In *Encyclopedia of Education* (2nd ed., pp. 1370-1373). New York, NY: Macmillan Reference.

Schenker, A. (2003). *Graph-theoretic techniques for web content mining* [Unpublished doctoral dissertation]. University of South Florida.

Schwab, K. (2016). *The fourth industrial revolution*. New York, NY: Crown Publishing Group.

Scott, J. (1988). Social network analysis. *Sociology, 22*(1), 109-127. doi:10.1177/0038038588022001007

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380-1400. doi:10.1177%2F0002764213498851

Tunstall, P., & Gsipps, C. (1996). Teacher feedback to young children in formative assessment: A Typology. *British Educational Research Journal, 22*(4), 389-404. doi:10.1080/0141192960220402

Van den Bergh, L., Ros, A., & Beijaard, D. (2012). Teacher feedback during active learning: Current practices in primary schools. *British Journal of Educational Psychology, 83*(2), 341-362 doi:10.1111/j.2044-8279.2012.02073.x

Williamson, B. (2019). Digital policy sociology: Software and science in data-intensive precision education. *Critical Studies in Education*, 1-17 doi:10.1080/17508487.2019.1691030

Wollenschläger, M., Hattie, J., Machts, N., Möller, J., & Harms, U. (2016). What makes rubrics effective in teacher-feedback? Transparency of learning goals is not enough. *Contemporary Educational Psychology, 44*, 1-11. doi:10.1016/j.cedpsych.2015.11.003

Wortham, D. W. (1999). Nodal and matrix analyses of communication patterns in small groups. In Hoadley, C. M. & Roschelle, J. (Eds.), *Proceedings of the Computer Support for Collaborative Learning (CSCL) 1999 Conference* (pp. 681-686). Palo Alto, CA: International Society of the Learning Sciences. https://repository.isls.org//handle/1/4382

Yang, S. J. H. (2019). Precision education: New challenges for AI in education [conference keynote]. In *Proceedings of the 27th International Conference on Computers in Education (ICCE)* (pp. XXVII-XXVIII). Kenting, Taiwan: Asia-Pacific Society for Computers in Education (APSCE).

Zheng, Y., & Yu, S. (2018). Student engagement with teacher written corrective feedback in EFL writing: A Case study of Chinese lower-proficiency students. *Assessing Writing, 37*, 13-24. doi:10.1016/j.asw.2018.03.001

Luan, H., & Tsai, C.-C. (2021). A Review of Using Machine Learning Approaches for Precision Education. *Educational Technology & Society*, *24* (1), 250–266.

A Review of Using Machine Learning Approaches for Precision Education

Hui Luan[1] and Chin-Chung Tsai[1, 2*]
[1]Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, Taiwan //
[2]Program of Learning Sciences, National Taiwan Normal University, Taiwan // hluanv@gmail.com //
tsaicc@ntnu.edu.tw
*Corresponding author

ABSTRACT: In recent years, in the field of education, there has been a clear progressive trend toward precision education. As a rapidly evolving AI technique, machine learning is viewed as an important means to realize it. In this paper, we systematically review 40 empirical studies regarding machine-learning-based precision education. The results showed that the majority of studies focused on the prediction of learning performance or dropouts, and were carried out in online or blended learning environments among university students majoring in computer science or STEM, whereas the data sources were divergent. The commonly used machine learning algorithms, evaluation methods, and validation approaches are presented. The emerging issues and future directions are discussed accordingly.

Keywords: Precision education, Personalized learning, Individualized learning, Machine learning, Individual differences

1. Introduction

A key goal of education is to cultivate the talents of all students. It is commonly believed that each student's learning experiences are unique. Therefore, it is imperative to teach in line with each individual's ability or rhythm (Corno & Snow, 1986). However, in the traditional education paradigm, the one-size-fits-all approach is adopted with a focus on average students (Cook et al., 2018). It is almost impossible for the teacher to implement tailor-made pedagogical tools to cater to students' diverse learning styles and needs.

Personalized learning refers to "instruction in which the pace of learning and the instructional approach are optimized for the needs of each learner. Learning objectives, instructional approaches, and instructional content (and its sequencing) may all vary based on learner needs. In addition, learning activities are meaningful and relevant to learners, driven by their interests, and often self-initiated" (U.S. Department of Education, 2017, p. 9). Educational technologies in personalized learning (e.g., e-learning system adaptive to learners' learning style, knowledge level, interest, and preference) help students to learn more effectively than pure educational psychology theories or methods (e.g., Huang et al., 2012; Klašnja-Milićević et al., 2011; Tseng et al., 2008). To date, a large body of these personalized systems used traditional computers or devices. In contrast, smart devices such as wearable devices, smartphones, and tablet computers were less frequently used, and artificial intelligence has a significant impact on these personalized learning systems (Xie et al., 2019).

In the field of education, with the rapid advances in artificial intelligence and data science, accurate and rich learning data are able to be collected and to reveal learning patterns and specific learning needs. Accordingly, an "optimal" personalized learning path or feedback can be provided. As a result, there is a clear progressive shift from a one-size-fits-all approach to precision education (Lu et al., 2018; Tsai et al., 2020). Precision education considers the individual differences of learners in their learning environments, identifies at-risk students as early as possible and provides timely and tailoring intervention accompanied with proper teaching materials and strategies, and learning strategies and activities (Cook et al., 2018; Frey, 2019; Lu et al., 2018). Mirroring precision medicine (Collins & Varmus, 2015), described as "an innovative approach to disease prevention and treatment that takes into account individual differences in people's genes, environments, and lifestyles" (The White House, 2015), precision education aims to advance the diagnosis, prediction, treatment, and prevention of at-risk students (Yang, 2019). According to Hart (2016), the more immediate goals of precision education are to obtain an accurate understanding of the learner's unique individual needs through profiling or diagnosis. In contrast, its ultimate goals are to implement individualized treatment or prevention and to enhance individual student's learning outcomes. Regarding the means to achieve its goals, precision education stresses the importance of advanced computational technologies such as learning analytics, artificial intelligence, and machine learning (Williamson, 2019; Yang, 2019).

Situated at the core of AI and data science, machine learning is one of the most rapidly growing techniques and has come to be expected as an essential means to achieve precision education and optimize learning. Machine learning addresses the question of how to construct computer systems that can learn automatically from past experiences without explicit programming (Jordan & Mitchell, 2015). Machine learning algorithms can classify profiling and patterns, provide new models and insights, and make predictions and recommendations to customize each individual's needs and circumstances. With the availability of adequate training data and low-cost data analytics tools, the data-intensive machine learning methods are widely used to facilitate evidence-based decision making in commerce, medicine, science, agriculture, and manufacturing (Kourou et al., 2015; Liakos et al., 2018; Lin et al., 2011; Ramprasad et al., 2017; Voyant et al., 2017; Wu et al., 2017). Machine learning has attracted growing interest in education recently (Gobert & Sao Pedro, 2017; Zhou et al., 2018). By adopting machine learning in education, the learning paths can be changed dynamically and personalized based on the learner's progress and pace (Kuch et al., 2020). Therefore, individualized learning, which is adaptive to individual needs in real-time, has recently gained increasing attention from educational researchers (Lu et al., 2018).

The ideas of "personalized learning," "individualized learning," and "precision education" are often interchangeable, while the concept of precision education is relatively new. The term "precision education" first appeared in 2016 (Hart, 2016). Since then, a growing number of studies in this research area have adopted machine learning methods. The application of machine learning in education has been reviewed in science assessment (Zhai et al., 2020) and educational technology (Korkmaz & Correia, 2019). Nevertheless, a systematic review of the applications of machine learning in precision education is lacking. The development, trends, and challenges of technology-supported adaptive/personalized learning have recently been systematically reviewed (Xie et al., 2019). However, Xie et al. (2019) found that one study adopted deep neural network techniques among the 70 studies published from 2007 to 2017 (Shi & Weninger, 2017). To provide insights into the educational benefits of machine learning, a comprehensive review is needed to fill the gaps and shed light on the current status, major challenges, and potential future directions of using a machine learning approach for precision education. Five specific research questions guide this study:

- What are the primary research purposes of using a machine learning approach for precision education (e.g., diagnosis, prediction, treatment, or prevention)?
- In which learning environments, domains, education levels, samples, and data sources have machine learning been applied for precision education?
- What are the learner's individual differences and learning outcomes of using machine learning in precision education?
- What are the algorithms, evaluation measures, and validation approaches of using machine learning in precision education?
- Are there significant relationships among these aforementioned categorical variables by using chi-square analysis?

2. Methods

2.1. Literature search

For this review, we selected peer-review articles employing machine learning techniques for precision education, published in journals that are indexed in the Web of Science database. This is a highly reputable database in terms of research quality. As machine learning and precision education have become popular since 2016, we set our search to articles published from January 2016 to July 2020. We further limited the document type to journal or early access articles written in English to ensure the consistent quality of the recruited studies. The keywords used for this search included "machine learn*," "machine-learn*," "precision," "personal* (not personality)," "individual*," and "education." Based on the above search parameters, a total of 151 articles were retrieved. Figure 1 demonstrated the process of selecting the eligible studies for this review. We then screened the articles by reading the titles and abstracts. Those studies that matched our inclusion criteria were retained. The inclusion criteria are fourfold: (1) empirical studies (not a position paper or review paper), (2) in a learning setting, (3) using machine learning techniques, and (4) measuring individual differences. When one of the inclusion criteria was not met, the study was excluded. For instance, Koutsouleris et al. (2016) adopted a machine learning approach to predict the treatment outcomes in patients with first-episode psychosis. This study was excluded because it belongs to the field of psychiatry. By the end of this stage, a total of 34 studies were eligible, while 22 studies were uncertain as the information provided in the title and abstract was insufficient to make a judgment. We then reviewed all of the full-text articles for the uncertain studies. After applying our selection criteria, the final dataset comprised 40 empirical studies.

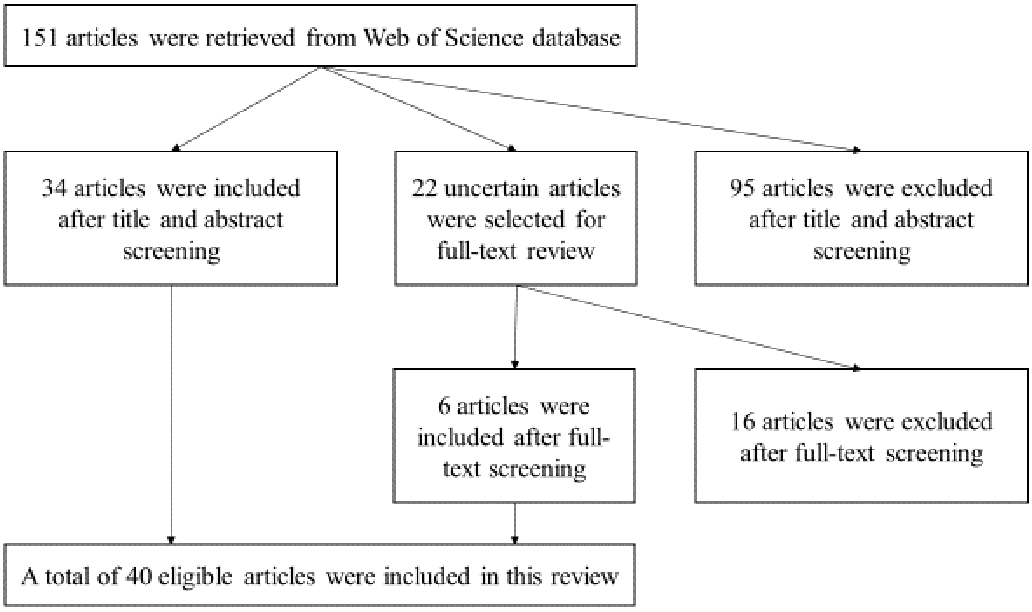


Figure 1. The selection of eligible studies

2.2. Coding scheme

To analyze the current status and future trends of machine-learning-based precision education, all studies were qualitatively coded. The coding scheme consists of eight main categories, including research purpose, general information, individual differences, learning outcomes, machine learning algorithms, evaluation of algorithms, validation of algorithms, and major research findings. Table 1 illustrates the overall coding scheme of this study.

As suggested by Yang (2019), the coding for the research purpose includes diagnosis or profiling (e.g., introverts, extroverts), prediction (e.g., dropout, performance), treatment or intervention (e.g., plan-making intervention, value-relevance intervention), prevention, and recommendations (e.g., personalized learning paths, learning contents).

The coding for general information comprises publication year, learning environment, learning domain, learners' educational level, sample size, data source. The coding for the learning environment can be further divided into classroom, online, blended (combining classroom and online learning; Graham, 2006), and others (e.g., laboratory). The coding for the learning domain is grouped into four main categories, including one multiple domains and three single domains. That is, multiple domains with many disciplines in a single study, computer sciences domain (e.g., programming, internet of things, data science), STEM domain (e.g., engineering, mathematics, digital design, electronic technology), and social sciences domain (e.g., finance, statistics, psychology, and language learning). Note that the main reason for selecting computer sciences as a separate category is that computer sciences are very closely linked with precision education in terms of that precision education places great emphasis on data-intensive digital technologies (Williamson, 2019). In addition, the STEM domain, an interdisciplinary domain that encompasses disciplines of science, technology, engineering, and mathematics, is categorized as a single primary domain in this study for two reasons. The first is that STEM education has gained increasing prominence in the past decade (Honey et al., 2014). The second is that STEM topics are often step-based and well-defined problems, where artificial intelligence tools and techniques can be relatively easily applied (Humble & Mozelius, 2019; Roll & Wylie, 2016). The coding for learners' educational level includes K-12 students, university students, and others (e.g., teachers, working adults). The coding for sample size is classified into 1-999, 1,000-9,999, more than 10,000, and others (e.g., number of responses in training dataset). The coding for the data source is divided into four major categories, including log files from a learning platform (e.g., MOOCs, e-learning system, Facebook, Mobile app), learning records or surveys (e.g., prior grades, past performance, satisfaction), institutional database (e.g., SAT scores, financial status), and physiological records (e.g., EEG signals, eye-movement data).

The coding for individual differences consists of seven categories: demographic (e.g., gender, age), academic (e.g., past performance, prior knowledge), cognitive (e.g., reasoning, working memory), affective (e.g., self-concept, motivation, learning styles, relationships with teachers and peers), behavioral (e.g., log activities, time and efforts for learning activities), pedagogical. In particular, pedagogical category refers to the factors relevant to course difficulty, learning content, or classroom characteristics. The coding for learning outcomes includes performance and dropout/ attrition/retention. If a learning outcome could not be categorized into these categories, the coder tagged it as others and wrote down the specific information.

Based on Moreno-Marcos et al.'s (2018) and Zhai et al.'s (2020) review studies, the coding for machine learning algorithms includes the commonly used ones such as K-Nearest Neighbors (KNN), Naïve Bayes, Regression, Random Forest, Decision Tree, Support Vector Machine (SVM), and Neural Networks. The relatively infrequent algorithms are categorized into the "others" category. This classification and the definitions of each type of algorithm evaluation and validation approaches were adopted from Zhai et al.'s (2020) review study on the applications of machine learning methods in science assessment. The coding for evaluation measure of algorithms covers the most frequently used indicators, including accuracy, precision, recall/sensitivity, F1-score, Area under the ROC Curve (AUC), and Receiver Operating Characteristic (ROC). The less often used evaluation indicators are classified as others. In this review, the validation approaches of algorithms are classified into self-validation, split validation, and cross-validation.

Table 1. Coding scheme of this study

Coding categories	
1. Research Purpose Diagnosis (e.g., introverts, extroverts) Prediction (e.g., dropout, performance) Intervention (e.g., plan-making intervention, value-relevance intervention) Prevention Recommendations (e.g., personalized learning paths, learning contents) **2. General information** Publication year Learning environment • Classroom • Online • Blended • Others Learning domain • Computer sciences • STEM • Social sciences • Multiple Learners' education level • K-12 students • University students • Others Sample size • 1-999 • 1,000-9,999 • >=10,000 • Others Data source • Log files from learning platform (e.g., log activities in MOOCs) • Learning records or surveys (e.g., prior grades, satisfaction) • Institutional databases (e.g., SAT scores, financial status) • Physiological records (e.g., EEG signals, eye-movement data)	**3. Individual differences** Demographic (e.g., gender, age) Academic (e.g., past performance, prior knowledge) Cognitive (e.g., reasoning, working memory) Affective (e.g., self-concept, motivation, learning styles, relationships with teachers and peers) Behavioral (e.g., log activities, time and efforts for learning activities) Pedagogical (e.g., learning contents, course difficulty) **4. Learning outcomes** Performance Dropout/ attrition/retention Others **5. Machine learning algorithms** KNN Naïve Bayes Regression Random Forest Decision Tree SVM Neural Networks Others **6. Evaluation of algorithms** Accuracy Precision Recall/sensitivity F1-score AUC ROC Others **7. Validation of algorithms** Self-validation Split validation Cross-validation **8. Major research findings**

Finally, among the coding scheme, individual differences, machine learning algorithms, evaluation of algorithms, and validation of algorithms are coded with multiple responses as a single study generally measures several individual characteristics (e.g., demographic, academic, behavioral) and adopted more than one algorithm, evaluation measures, and validation approach. As a result, the sum of these factors may exceed 40. In contrast, the rest of the coding categories are multiple-choice items, while the sums are all 40.

3. Results

In this section, the results of the analyses of the 40 empirical studies are presented. By and large, the distribution of the publication year clearly indicated that there is a growing trend. There was only one paper published in 2016, two in 2017, and three in 2018. In the year of 2019, there were 17 studies published. This year was obviously the turning point, as there were 17 articles employing machine learning techniques for precision education within the first seven months of 2020. This result corroborates the rising popularity of this topic. These 40 reviewed studies were analyzed following five major research questions regarding the research purpose, general information, learners' characteristics and learning outcomes, and machine learning algorithms and algorithms' evaluation. Each research question is in its own subsection. In addition to descriptive analysis, Chi-square analysis was also conducted to examine whether there are statistically significant relationships among the aforementioned categorical variables with multiple-choice items. The significant results are reported accordingly.

3.1. What are the primary research purposes of using a machine learning approach for precision education?

As shown in Figure 2, the analyses indicated that among the 40 studies, the majority (25 studies) adopted machine learning methods to make predictions, and nine aimed for diagnosis and profiling. Only three studies carried out intervention research such as using behavioral intervention (e.g., self-regulation, value-relevance; Kizilcec et al., 2020) and machine learning-generated individual eye movement feedback (Krol & Krol, 2019). It was noted that no study employed prevention. Furthermore, three studies provided recommendations such as appropriate learning contents through an individualized AI tutor (Kim & Kim, 2020). Among these 40 studies, only two adopted an experimental design (Kizilcec et al., 2020; Krol & Krol, 2019), and two adopted a quasi-experimental design (Magana et al., 2019; Ninaus et al., 2019).

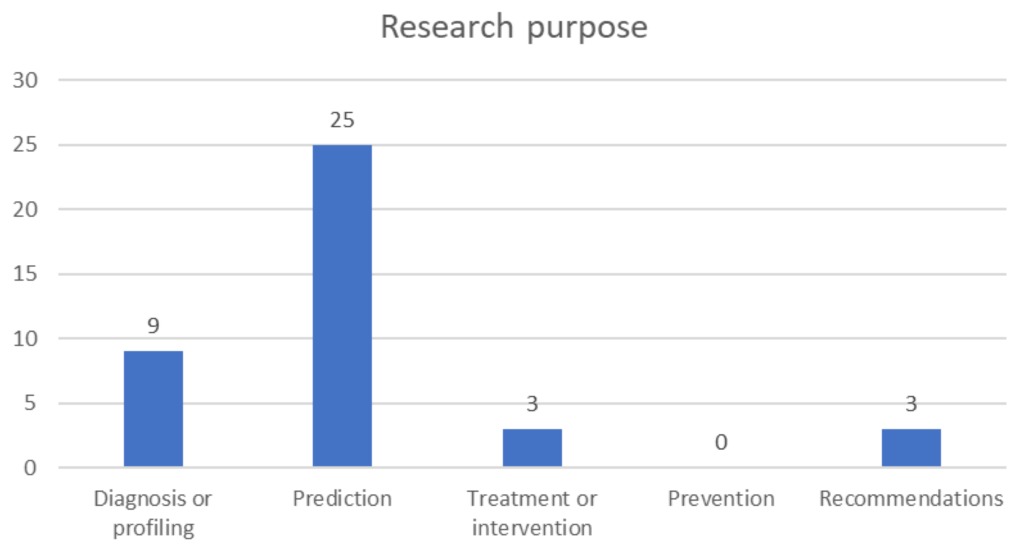


Figure 2. Distribution of research purposes

3.2. In which learning environments, domains, education levels, samples, and data sources have machine learning been applied for precision education?

In terms of the learning environments in which machine learning was employed, four categories were defined: classroom, online, blended, and others (Figure 3). The most used learning environment was online learning (18 out of 40 studies), followed by classroom learning (9 studies) and blended learning (8 studies). The studies not specified, and the study conducted in a game-based learning environment were tagged as others (5 studies).

With regard to the learning domain, most studies recruited learners from multiple domains (n = 10) partially due to the large dataset of these studies (Figure 4). Concerning specific learning domains, STEM (n = 10; e.g., engineering, mathematics, digital design, electronic technology) and computer science (n = 9; e.g., programming, internet of things, data science) related domains were most popular. Machine learning was also used in social sciences (n = 7) such as finance, statistics, psychology, and language learning. Besides, there were four studies which did not specify the learning domain.

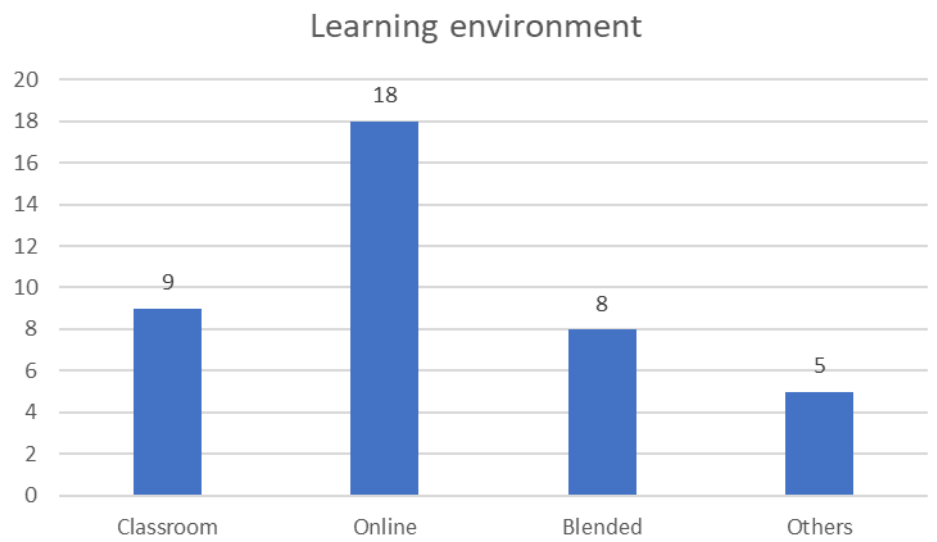


Figure 3. Distribution of the learning environment

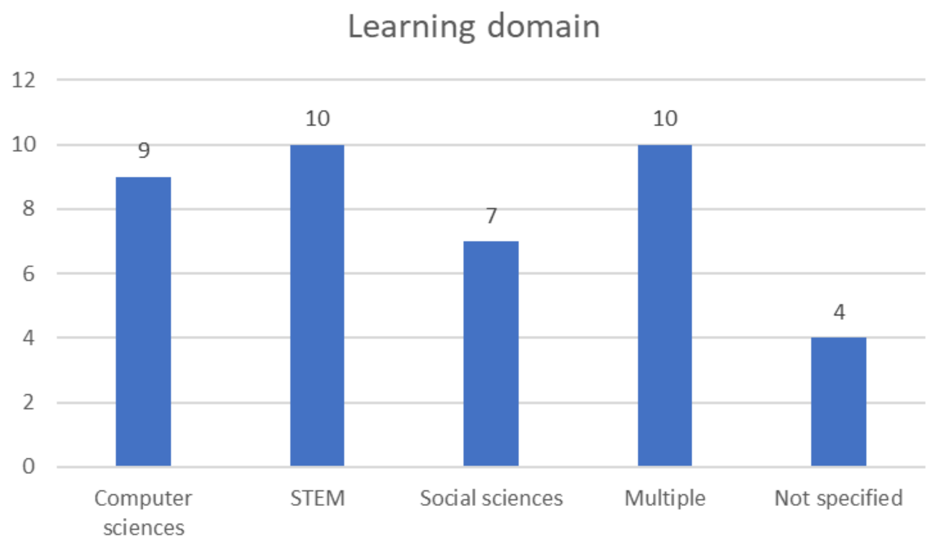


Figure 4. Distribution of the learning domain

As shown in Figure 5, the overwhelming majority of studies recruited university students (24 out of 40 studies), whereas five focused on elementary students and secondary students. Among the 40 reviewed studies, nine were tagged as others. The reason is that these studies recruited participants from heterogeneous groups by using MOOCs or other large databases. Among the nine studies categorized as others, one solicited participants from teachers who were taking a computer science course (Spatiotis et al., 2020). Besides, there were two studies which did not specify the learners' education level.

To train a good machine learning algorithm model, a certain amount of data is required. Of the 40 studies included in this paper, the number of students ranged from 6 to 269,169. Among them seven studies' sample size was beyond 10,000, and six studies recruited 1,000-9,999 students. Surprisingly, the sample size of most of the studies (22 studies) fell into the 1-999 range. There were three studies which reported the numbers of responses for the training data instead, namely 3,000, 11,156, and 76,936, respectively. The training responses ranged from 650 to 9,966,292. For the study which recruited six students (Kurilovas, 2018), the responses used for training data were 5,658. There were three studies found with no sample size or training response information. Figure 6 illustrates the distribution of sample size.

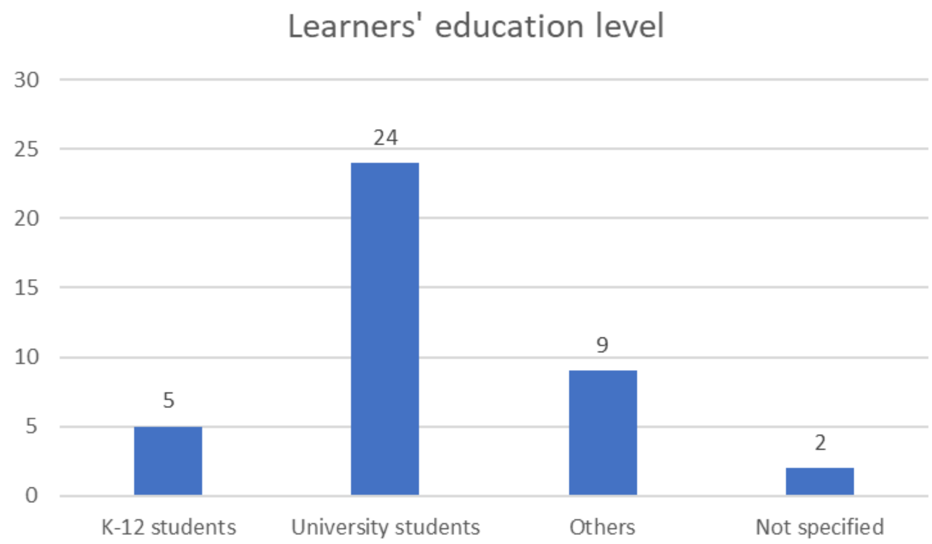


Figure 5. Distribution of learners' education level

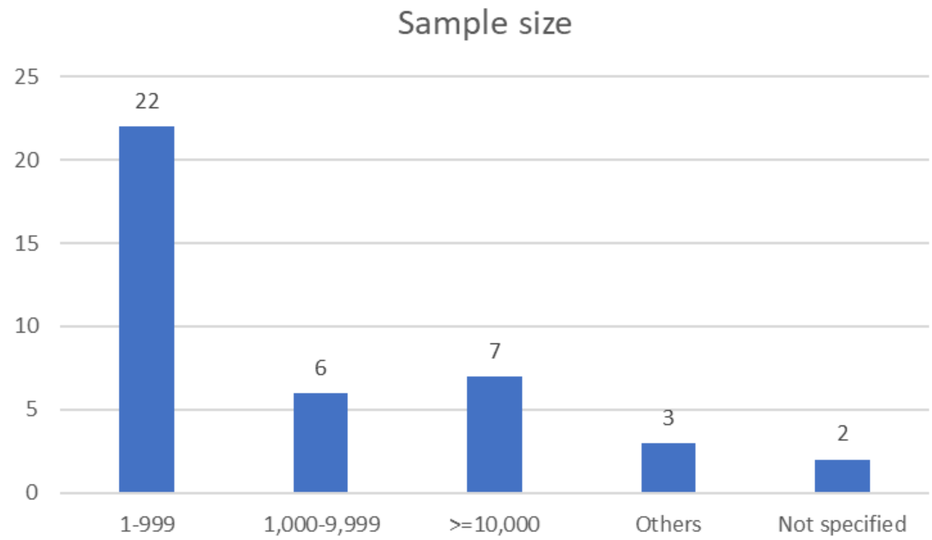


Figure 6. Distribution of sample size

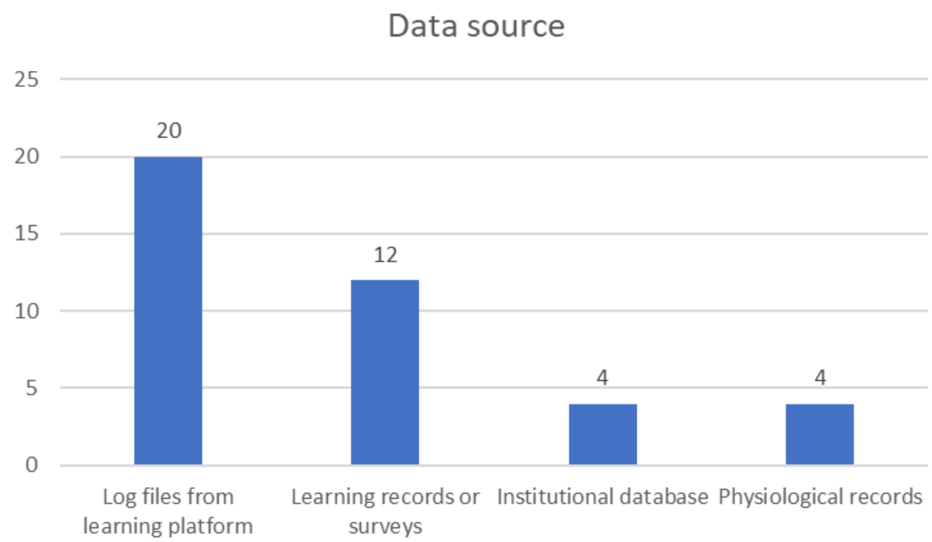


Figure 7. Distribution of data sources

As shown in Figure 7, the distribution of major data sources varied. Most of the studies collected students' log files from a learning platform (20 studies; e.g., MOOCs, e-learning system, Facebook, Mobile app), followed by learning records or surveys (12 studies; e.g., prior grades, past performance, satisfaction) and institutional databases (4 studies). It is worth noting that several neuroscience studies began to adopt machine learning to analyze the physiological records (4 studies). In particular, there were two studies which collected data from EEG signals (Luo & Zhou, 2020; Rajkumar & Ganapathy, 2020), and two which analyzed eye-movement data employing a machine learning approach (Krol & Krol, 2019; Lee et al., 2019).

3.3. What are the learners' individual differences and learning outcomes of using machine learning in precision education?

For the machine learning training model's features, the individual characteristics were grouped into six main categories as described in the coding scheme section (Figure 8). Among them, the behavioral characteristics (e.g., log activities, time and efforts for learning activities) were mostly investigated (18 studies), followed by affective factors (16 studies). The affective factor included self-concept, motivation, attitudes, learning styles, learning strategies, coping strategies, relationships with teachers and peers, time management, etc. For example, Rajkumar and Ganapathy (2020) compared Chatbot and machine learning algorithms' classification accuracy. Based on the VARK (Visual, Auditory, Read/Write and Kinesthetic) learning style, individuals were classified as Introverts or Extraverts. Introverts prefer to study alone in calm places. Extraverts prefer to study in a group and like to learn with music and audiobooks. Participants first answered questionnaire questions via the Chatbot, then their learning beta brain waves were recorded through a non-invasive EEG sensor while they were processing visual and auditory learning content. The result showed that the classification accuracy of the Chatbot and machine learning algorithms was similar, whereas the Chatbot was fast and convenient. There were 14 studies which used demographic information and 13 which examined the academic characteristics (e.g., past performance, prior knowledge). There were six studies which involved cognitive factors (e.g., reasoning, working memory) as independent variables in the training model. Finally, among the 40 studies, five examined pedagogical factors relevant to course difficulty or classroom characteristics, such as learning materials (e.g., Coussement et al., 2020; Kassak et al., 2016), teaching strategies or interventions (Kizilcec et al., 2020).

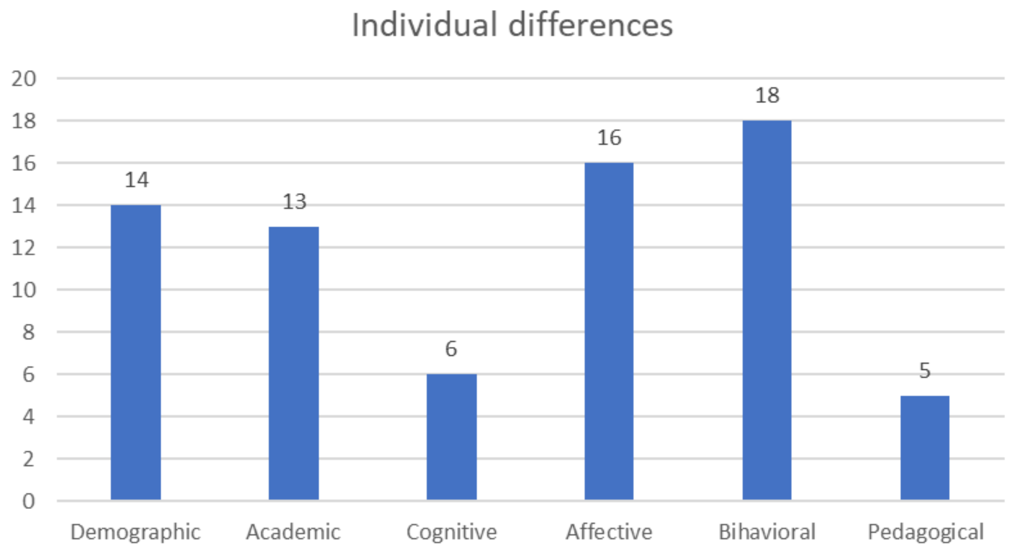


Figure 8. Distribution of individual differences

Investigating the learning outcomes, as shown in Figure 9, the majority of studies focused on learning performance (15 studies) and dropout/attrition/retention (11 studies). This finding was consistent with the research purpose, as most of the studies aimed to make predictions or profiling and tried to identify at-risk students as early as possible. Of the 40 reviewed studies, nine were categorized as others, such as emotion engagement (Ninaus et al., 2019), satisfaction (Spatiotis et al., 2020), decision quality (Krol & Krol, 2019), behavior intentions (Arpaci, 2019), and brain organization (Astle et al., 2019). For instance, using an artificial neural network algorithm named Self Organising Maps, Astle et al. (2019) grouped 530 heterogeneous struggling children into four clusters with distinctive cognitive profiles. The group comparison showed that they were significantly different in terms of their learning performance (e.g., reading, math), behavioral scores (e.g., executive function, communication), and patterns of brain organization. Besides, Ninaus et al. (2019) examined emotional engagement differences in game-based and non-game-based learning by using the SVM algorithm. There were five studies which were tagged as not applicable.

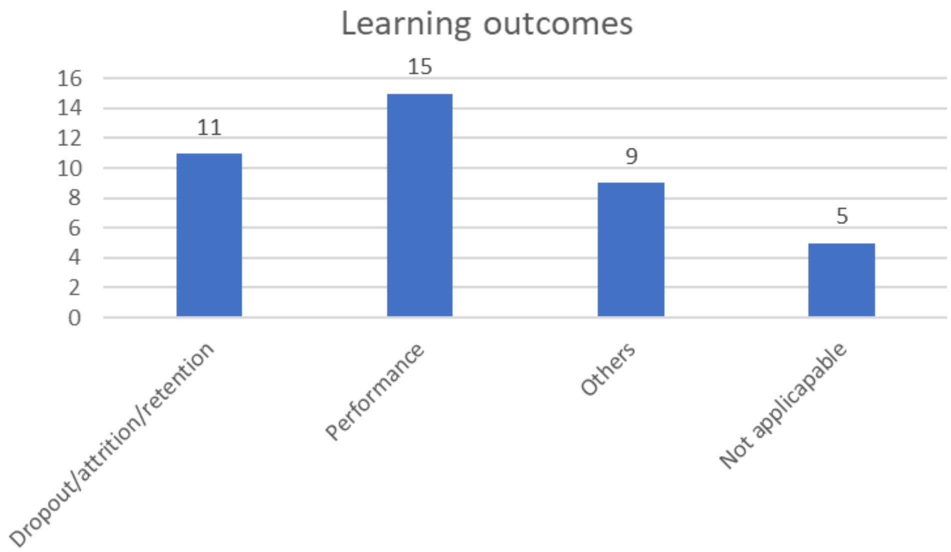


Figure 9. Distribution of learning outcomes

3.4. What are the algorithms, evaluation measures, and validation approaches for using machine learning in precision education?

The machine learning experts developed different kinds of algorithms to train the models. Depending on the data characteristics, selecting some algorithms may result in better performance. Various indicators can evaluate the quality of the model according to the research purpose. Also, in machine learning, whether a trained model can be generalized to an independent dataset like a testing dataset is essential. This process refers to model validation. Investigating these reviewed studies, the commonly used algorithms were KNN, Naïve Bayes, regression, random forest, neural networks, decision tree, and SVM. In contrast, the innovative or uncommon machine learning algorithms were numerous (see Figure 10).

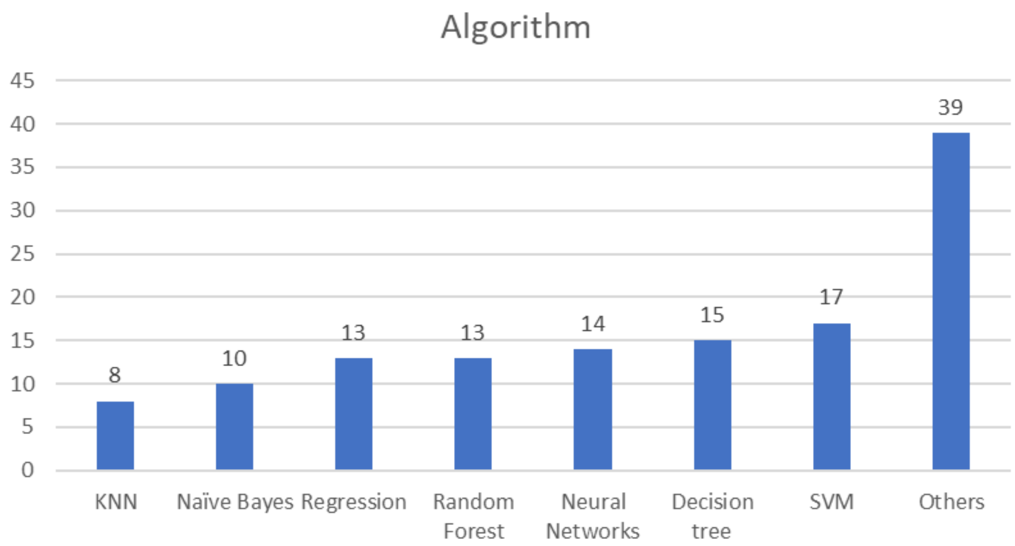


Figure 10. Distribution of machine learning algorithms

As shown in Figure 11, the models were often evaluated by using accuracy, precision, recall/sensitivity, F1-score, AUC, and ROC, while other measures were also used, such as specificity, PR curve, and RMSE. There were two studies which did not specify the algorithm used in their study.

As shown in Figure 12, adopting Zhai et al.'s (2020) classification and definitions, three validation approaches were classified in this review: self-validation, split validation, and cross-validation. Self-validation did not divide data into a training set and test set; in contrast, the same data were used to build the algorithmic model and to evaluate the model. Split validation divided data into two sets: a training dataset to train the model, and a testing dataset to evaluate the model. The generalizability of the algorithm was improved by using split validation.

However, there were only two datasets, and the validation indicator still may vary when the settings of the training and testing data were changed. Therefore, in machine learning, cross-validation is more commonly used. Cross-validation divides data into n subsets (n-fold; the number of n might be 4, 5, or 10), while with this process, each subset was used to be both the training set and the testing set. Figure 12 demonstrates that cross-validation was most frequently used in our review sample, too; that is, 24 out of 40 studies used cross-validation, followed by self-validation (16 studies) and split validation (10 studies).

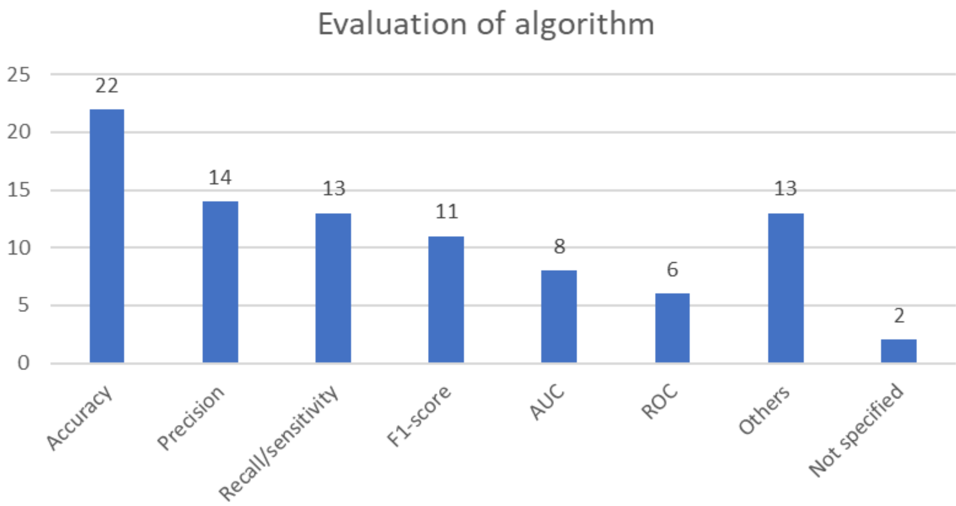


Figure 11. Distribution of machine learning algorithm evaluation

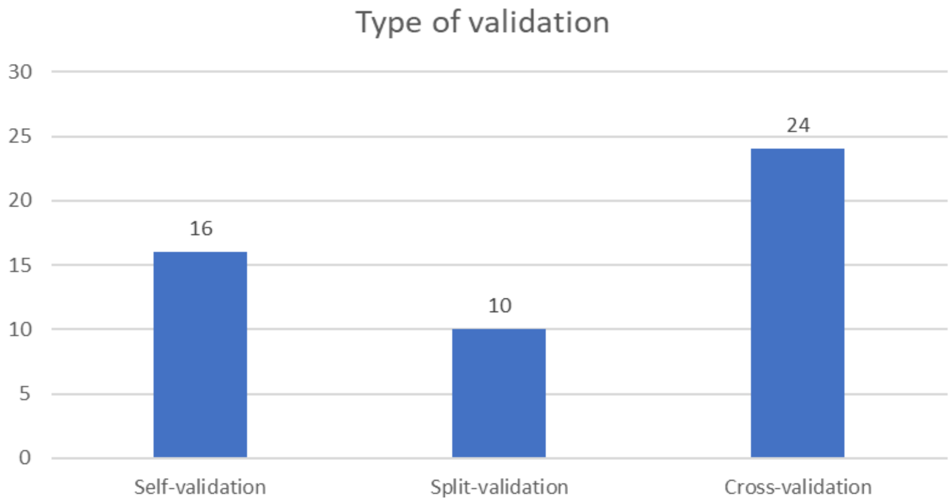


Figure 12. Distribution of validation of machine learning algorithms

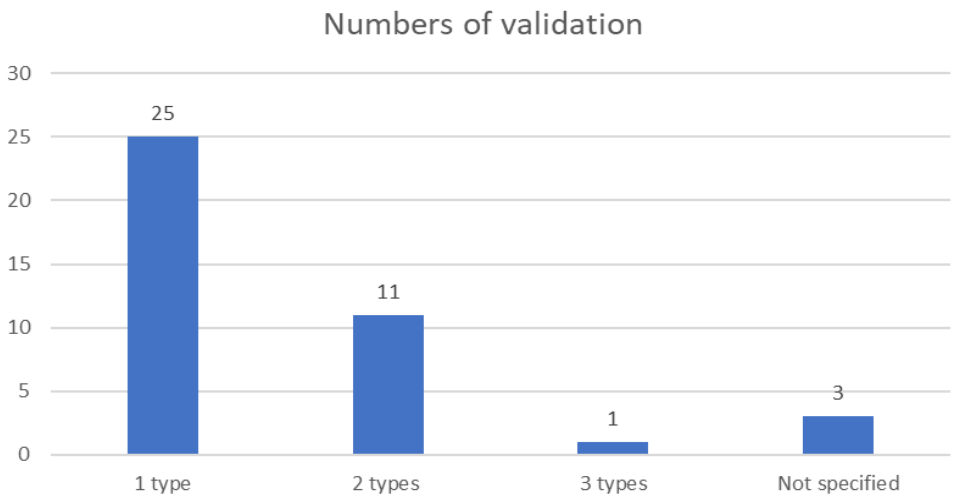


Figure 13. Distribution of validation types of machine learning algorithms

Figure 13 further presents how many types of validation approaches were used in our reviewed studies. The result showed that a single method was dominant (26 studies), followed by two mixed validation approaches (10 studies), and three types of validation used together (1 studies). Among the 40 reviewed studies, three did not report the validation approach.

3.5. Are there significant relationships among these aforementioned categorical variables by using chi-square analysis?

The Chi-square analysis showed that for the studies' sample size larger than 10,000, the data were mainly collected in online environments, whereas for the studies with sample sizes ranging from 1 to 999, the data were collected through classroom, online, or blended environments ($\chi^2 = 22.44$, $p = .033$). For the studies focused on a single learning domain (e.g., computer science), the data source is mainly from log files from the learning platform and learning records, while the studies which covered multiple learning domains also obtained data from institutional databases ($\chi^2 = 22.69$, $p = .031$).

Concerning learning outcomes, learning performance and dropout/attrition/retention were mostly measured by the studies aimed to make predictions, while learning outcomes were less often measured for the studies mainly focused on profiling or recommendations ($\chi^2 = 24.32$, $p = .004$). Furthermore, learning performance is equally important among all learning environments, while "other" learning outcomes are often measured in "other" learning environments ($\chi^2 = 28.07$, $p = .001$); for instance, emotional engagement was measured in game-based learning environments (Ninaus et al., 2019), or brain organization was evaluated by laboratory study (Astle et al., 2019).

With respect to the validation of algorithm, self-validation and split-validation were seldom used together ($\chi^2 = 6.07$, $p = .013$), whereas self-validation was more often combined with cross-validation ($\chi^2 = 5.52$, $p = .019$). Cross-validation tent to be used as the only validation approach, followed by combination with another type of validation (e.g., self-validation, split-validation; $\chi^2 = 9.62$, $p = .008$). Split-validation were often paired with cross-validation, followed by using alone ($\chi^2 = 6.14$, $p = .047$).

4. Discussion

In this paper, we systematically reviewed the emerging field of using a machine learning approach for precision education. After a series of screening steps, a total of 40 studies remained. To date, profiling and prediction were the primary research purpose. In short, most studies were carried out in an online or blended learning environment among university students majoring in computer science or STEM with heterogeneous data sources, such as MOOCs, institutional datasets, learning records, etc. The results indicated that using a machine learning approach for precision education is a fast-growing area with high potential. The emerging issues and future directions are presented in the following section.

4.1. From personalized learning to individualized learning

In the traditional education setting, personalized learning is resource-heavy and time-consuming. With the development of society and technology, data-driven personalized learning or precision education has become an achievable education paradigm. Among the potential methodologies to realize precision education, machine learning is viewed as one of the most promising means that emphasizes individual-level support rather than class- or group-level assistance. Harnessing big data, machine learning approaches are capable of extracting meaningful patterns and making individualized predictions. Correspondingly, this review study showed that diagnosis and prediction were the most prevalent research types, consistent with the results from an earlier review study on artificial intelligence applications in higher education (Zawacki-Richter et al., 2019). Only a handful of studies ($n = 3$) delivered interventions, similar to the proportion of review studies adopting learning analytics in higher education (Sonderlund et al., 2019; Viberg et al., 2018). There was as yet no study that provided prevention. The rapid advances in automatic emotion detection techniques are opening up new possibilities to monitor students' real-time emotions (Ninaus et al., 2019) and to provide immediate individualized feedback or learning materials. Real-time feedback or interventions are encouraged in future studies to realize the machine learning technique's full potential. That is, the research focus may shift from personalized learning to individualized learning (Luan et al., 2020).

4.2. Domain and population generalization issues

To date, most studies have been carried out among university students majoring in computer sciences and STEM. It is reasonable that the researchers with information technology expertise and in-depth domain knowledge were more familiar with machine learning techniques and gained greater access to students in these domains. On the other hand, the step-based and well-defined problems in computer sciences and STEM topics were more likely for machine learning researchers to design and implement personalized educational tools or systems (Humble & Mozelius, 2019; Roll & Wylie, 2016). Since the number of published articles in the first seven months of the year of 2020 was equal to the total of the year of 2019, the adoption of machine learning in education might experience a growth spurt, as occurred with AI-based precision medicine (Kourou et al., 2015; Krittanawong et al., 2017; Rajkomar et al., 2019) and precision psychiatry (Bzdok & Meyer-Lindenberg, 2018). Perhaps future research could broaden the learning domain from computer sciences and STEM to a more general knowledge area and include more learners at lower education levels such as kindergartens, elementary and secondary students.

4.3. Convergence of machine learning and neuroscience

The other notable line in research employing machine learning methods for precision education is the convergence of machine learning and neuroscience, similar to the existing trends in psychiatry (Janssen et al., 2018). The vast amount of data generated by EEG and Eye-movement devices is a perfect match for machine learning. The algorithmic models can be utilized to classify the patterns of cognitive ability such as working memory (Luo & Zhou, 2020) and styles of attention in financial decision making (Krol & Krol, 2019). Real-time feedback provided by machine learning techniques enables students to significantly improve their performance (Krol & Krol, 2019). The knowledge and insights from different forms of learning data are converging to create a new interdisciplinary science of learning that is capable to provide differentiated educational practices (Kuch et al., 2020).

4.4. Integration of innovative technologies and classic learning theories

Last but not least, we noticed that most of these reviewed studies selected features based on data availability; thus, past performance and log activities were frequently used in the training models. These data-intensive machine learning technologies might be integrated with learning theories to more effectively enhance students' learning (X. Chen et al., 2020; Hew et al., 2019). It is essential to help students become active participants in their own learning process and facilitate their self-directed learning (Loftus & Madden, 2020). Depending on the individual needs and specific educational purposes, pedagogical tools and learning strategies can be designed from different education perspectives such as behaviorism, information processing theory, social cognitive theory, and constructivism (Schunk, 2020).

5. Conclusion

In this review, we systematically reviewed 40 empirical studies regarding machine-learning-based precision education and showed that this field is a rapidly expanding area. This study uncovered the research gaps and provided an overview of the recent progress to help researchers understand essential topics in this emerging field. The results indicated that the majority of studies focused on the prediction of learning performance or dropouts, and were carried out in an online or blended learning environment among university students majoring in computer science or STEM, whereas the data sources were divergent and the sample size was 1-999. The commonly used machine learning algorithms, evaluation methods, and validation approaches were presented. This study offered valuable insights into the state-of-the-art machine learning techniques in precision education. We also discussed the emerging issues and critical directions to inspire the researchers interested in this field to conduct more empirical studies in the future. Furthermore, the research findings provided beneficial information for teachers and practitioners. The learning patterns and needs, and the predictions of learning outcomes generated by machine learning methods can help teachers make more precise decisions and reduce educational waste in time and resources.

This review study has several limitations. First, it should be noted that the current study mainly conducted a descriptive quantitative analysis of the current status of machine-learning-based precision education. Research synthesis and meta-analysis are recommended to provide more critical information. Second, we limited our search to journal articles to ensure the research quality, while book chapters and conference papers were excluded. Third, we set our search in published journal articles indexed in a highly reputable database, namely the Web of Science. Future studies can search for papers without these limitations to obtain more eligible items.

Compared to studies of other more mature educational technologies such as augmented reality and virtual reality (e.g., M. P. Chen et al., 2020; Cheng & Tsai, 2020; Jong et al., 2020), research using a machine learning approach for precision education is in its infancy. There is a long way to go in promoting learning and teaching by using machine learning methods. An in-depth understanding of the relationships between AI/machine learning techniques and an individual's characteristics calls for more subsequent research in this field.

Acknowledgement

This work was financially supported by the "Institute for Research Excellence in Learning Sciences" of the National Taiwan Normal University (NTNU) from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

References

(*Note.* References marked with stars (*) are the 40 reviewed studies in this paper)

*Abidi, S. M. R., Hussain, M., Xu, Y., & Zhang, W. (2019). Prediction of confusion attempting algebra homework in an intelligent tutoring system through machine learning techniques for educational sustainable development. *Sustainability*, *11*(1), 105. doi:10.3390/su11010105

*Arpaci, I. (2019). A Hybrid modeling approach for predicting the educational use of mobile cloud computing services in higher education. *Computers in Human Behavior, 90*, 181-187.

*Astle, D. E., Bathelt, J., Team, C., & Holmes, J. (2019). Remapping the cognitive and neural profiles of children who struggle at school. *Developmental Science, 22*(1), e12747. doi:10.1111/desc.12747

*Azcona, D., Hsiao, I. H., & Smeaton, A. F. (2019). Detecting students-at-risk in computer programming classes with learning analytics from students' digital footprints. *User Modeling and User-Adapted Interaction, 29*(4), 759-788.

Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine learning for precision psychiatry: Opportunities and challenges. *Biological Psychiatry-Cognitive Neuroscience and Neuroimaging*, *3*(3), 223–230.

Chen, M.-P., Wang, L.-C., Zou, D., Lin, S.-Y., Xie, H., & Tsai, C.-C. (2020). Effects of captions and English proficiency on learning effectiveness, motivation and attitude in augmented-reality-enhanced theme-based contextualized EFL learning. *Computer Assisted Language Learning*, 1-31.

Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *1*, 100002. doi:10.1016/j.caeai.2020.100002

Cheng, K. H., & Tsai, C. C. (2020). Students' motivational beliefs and strategies, perceived immersion and attitudes towards science learning with immersive virtual reality: A Partial least squares analysis. *British Journal of Educational Technology, 51*(6), 2139-2158.

Collins, F. S., & Varmus, H. (2015). A New initiative on precision medicine. *New England Journal of Medicine*, *372*(9), 793-795.

Cook, C. R., Kilgus, S. P., & Burns, M. K. (2018). Advancing the science and practice of precision education to enhance student outcomes. *Journal of School Psychology*, *66*(SI), 4–10.

Corno, L., & Snow, R. E. (1986). Adapting teaching to individual differences among learners. In *Handbook of research on teaching* (pp. 605-629). Chicago, IL: Rand McNally.

*Coussement, K., Phan, M., De Caigny, A., Benoit, D. F., & Raes, A. (2020). Predicting student dropout in subscription-based online learning environments: The Beneficial impact of the logit leaf model. *Decision Support Systems,* 113325. doi:10.1016/j.dss.2020.113325

*Delen, D., Topuz, K., & Eryarsoy, E. (2020). Development of a Bayesian belief network-based DSS for predicting and understanding freshmen student attrition. *European Journal of Operational Research, 281*(3), 575-587.

Frey, J. R. (2019). Assessment for special education: diagnosis and placement. *The ANNALS of the American Academy of Political and Social Science*, *683*(1), 149-161.

Gobert, J. D., & Sao Pedro, M. A. (2017). Digital assessment environments for scientific inquiry practices. In A. A. Rupp, & J. P. Leighton (Eds.), *The Wiley Handbook of Cognition and Assessment: Frameworks, Methodologies, and Applications* (pp. 508-534). West Sussex, UK: John Wiley & Sons Ltd.

*Gontzis, A. F., Kotsiantis, S., Panagiotakopoulos, C. T., & Verykios, V. S. (2019). A Predictive analytics framework as a countermeasure for attrition of students. *Interactive Learning Environments*, 1-15.

Graham, C. R. (2006). Blended learning systems. In *the Handbook of Blended Learning: Global Perspectives, Local Designs* (pp. 3-21). San Francisco, CA: Pfeiffer.

*Guerrero-Higueras, Á. M., Fernández Llamas, C., Sánchez González, L., Gutierrez Fernández, A., Esteban Costales, G., & Conde González, M. Á. (2020). Academic success assessment through Version Control Systems. *Applied Sciences, 10*(4), 1492. doi:10.3390/app10041492

Hart, S. A. (2016). Precision education initiative: Moving toward personalized education. *Mind Brain and Education*, *10*(4), 209–211. doi:10.1111/mbe.12109

Hew, K. F., Lan, M., Tang, Y., Jia, C., and Lo, C. K. (2019). Where is the "theory" within the field of educational technology research? *British Journal of Educational Technology*, *50*(3), 956–971.

Honey, M., Pearson, G., & Schweingruber, H. (Eds.). (2014). *STEM integration in K-12 education: Status, prospects, and an agenda for research*. Washington, DC: National Academies Press.

Huang, Y. M., Liang, T. H., Su, Y. N., & Chen, N. S. (2012). Empowering personalized learning with an interactive e-book learning system for elementary school students. *Educational Technology Research and Development*, *60*(4), 703-722.

Humble, N., & Mozelius, P. (2019). Artificial intelligence in education—A Promise, a threat or a hype? In *European Conference on the Impact of Artificial Intelligence and Robotics 2019 (ECIAIR 2019)* (pp. 149–156). Oxford, UK: Academic Conferences and Publishing International Limited.

*Hussain, M., Zhu, W., Zhang, W., Abidi, S. M. R., & Ali, S. (2018). Using machine learning to predict student difficulties from learning session data. *Artificial Intelligence Review, 52*(1), 381-407.

Janssen, R. J., Mourão-Miranda, J., & Schnack, H. G. (2018). Making individual prognoses in psychiatry using neuroimaging and machine learning. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(9),798-808. Doi: 10.1016/j.bpsc.2018.04.004

Jong, M. S. Y., Tsai, C. C., Xie, H., & Kwan-Kit Wong, F. (2020). Integrating interactive learner-immersed video-based virtual reality into learning and teaching of physical geography. *British Journal of Educational Technology*, *51*(6), 2063-2078.

Jordan, M. I., and Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255-260.

*Kassak, O., Kompan, M., & Bielikova, M. (2016). Student behavior in a web-based educational system: Exit intent prediction. *Engineering Applications of Artificial Intelligence*, *51*, 136-149.

*Kim, W.-H., & Kim, J.-H. (2020). Individualized AI tutor based on developmental learning networks. *IEEE Access, 8*, 27927-27937.

*Kizilcec, R. F., Reich, J., Yeomans, M., Dann, C., Brunskill, E., Lopez, G., Turkay, S., Williams, J. J., & Tingley, D. (2020). Scaling up behavioral science interventions in online education. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(26), 14900-14905.

Klašnja-Milićević, A., Vesin, B., Ivanović, M., & Budimac, Z. (2011). E-Learning personalization based on hybrid recommendation strategy and learning style identification. *Computers & Education*, *56*(3), 885-899.

Korkmaz, C., & Correia, A. P. (2019). A Review of research on machine learning in educational technology. *Educational Media International*, *56*(3), 250-267.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, *13*, 8–17.

Koutsouleris, N., Kahn, R. S., Chekroud, A. M., Leucht, S., Falkai, P., Wobrock, T., Derks, E. M., Fleischhacker, W. W., & Hasan, A. (2016). Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: A Machine learning approach. *The Lancet Psychiatry*, *3*(10), 935-946.

Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., & Kitai, T. (2017). Artificial intelligence in precision cardiovascular medicine. *Journal of the American College of Cardiology*, *69*(21), 2657–2664.

*Krol, M., & Krol, M. (2019). Learning from peers' eye movements in the absence of expert guidance: A Proof of concept using laboratory stock trading, eye tracking, and machine learning. *Cognitive Science, 43*(2), e12716. doi:10.1111/cogs.12716

Kuch, D., Kearnes, M., & Gulson, K. (2020). The Promise of precision: Datafication in medicine, agriculture and education. *Policy Studies*, 1-20.

*Kurilovas, E. (2018). Advanced machine learning approaches to personalise learning: learning analytics and decision making. *Behaviour & Information Technology, 38*(4), 410-421.

*Kushik, N., Yevtushenko, N., & Evtushenko, T. (2020). Novel machine learning technique for predicting teaching strategy effectiveness. *International Journal of Information Management, 53,*101488. doi:10.1016/j.ijinfomgt.2016.02.006

*Lee, S., & Chung, J. Y. (2019). The Machine learning-based dropout early warning system for improving the performance of dropout prediction. *Applied Sciences, 9*(15), 3093. doi:10.3390/app9153093

*Lee, S., Hwang, Y., Jin, Y., Ahn, S., & Park, J. (2019). Effects of individuality, education, and image on visual attention: Analyzing eye-tracking data using machine learning. *Journal of Eye Movement Research*, *12*(2), 4. doi:10.16910/jemr.12.2.4

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A Review. *Sensors*, *18*(8), 2674. doi:10.3390/s18082674

Lin, W. Y., Hu, Y. H., & Tsai, C. F. (2011). Machine learning in financial crisis prediction: A Survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *42*(4), 421-436.

*Loftus, M., & Madden, M. G. (2020). A Pedagogy of data and Artificial Intelligence for student subjectification. *Teaching in Higher Education, 25*(4), 456-475.

*López, M. B., Alor-Hernández, G., Sánchez-Cervantes, J. L., del Pilar Salas-Zárate, M., & Paredes-Valverde, M. A. (2018). EduRP: An educational resources platform based on opinion mining and semantic web. *Journal of Universal Computer Science*, *24*(11), 1515-1535.

Lu, O., Huang, A., Huang, J., Lin, A., Ogata, H., & Yang, S. J. H. (2018). Applying learning analytics for the early prediction of students' academic performance in blended learning. *Educational Technology & Society, 21*(2), 220-232.

Luan, H., Geczy, P., Lai, H., Gobert, J., Yang, S. J. H., Ogata, H., Baltes, J., Guerra, R., Li, P., & Tsai, C.-C. (2020). Challenges and future directions of big data and artificial intelligence in education. *Frontiers in Psychology*. *11,* 580820. doi:10.3389/fpsyg.2020.580820

*Luo, W., & Zhou, R. (2020). Can working memory task-related EEG biomarkers measure fluid intelligence and predict academic achievement in healthy children? *Frontiers in Behavioral Neuroscience*, *14*, 2. doi:10.3389/fnbeh.2020.00002

*Magana, A. J., Elluri, S., Dasgupta, C., Seah, Y. Y., Madamanchi, A., & Boutin, M. (2019). The Role of simulation-enabled design learning experiences on middle school students' self-generated inherence heuristics. *Journal of Science Education and Technology, 28*(4), 382-398.

Moreno-Marcos, P. M., Alario-Hoyos, C., Muñoz-Merino, P. J., & Kloos, C. D. (2018). Prediction in MOOCs: A Review and future research directions. *IEEE Transactions on Learning Technologies*, *12*(3), 384-401.

*Mubarak, A. A., Cao, H., & Zhang, W. (2020). Prediction of students' early dropout based on their interaction logs in online learning environment. *Interactive Learning Environments*, 1-20.

*Musso, M. F., Hernández, C. F. R., & Cascallar, E. C. (2020). Predicting key educational outcomes in academic trajectories: A Machine-learning approach. *Higher Education*, 1-20.

*Ninaus, M., Greipl, S., Kiili, K., Lindstedt, A., Huber, S., Klein, E., Karnath, H.-O., & Moeller, K. (2019). Increased emotional engagement in game-based learning – A Machine learning approach on facial emotion detection data. *Computers & Education, 142*, 103641–103610.

*Pliakos, K., Joo, S. H., Park, J. Y., Cornillie, F., Vens, C., & Van den Noortgate, W. (2019). Integrating machine learning into item response theory for addressing the cold start problem in adaptive learning systems. *Computers & Education*, *137*, 91-103.

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, *380*(14), 1347-1358.

*Rajkumar, R., & Ganapathy, V. (2020). Bio-inspiring learning style chatbot inventory using brain computing interface to increase the efficiency of e-learning. *IEEE Access, 8*, 67377-67395.

Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A., & Kim, C. (2017). Machine learning in materials informatics: Recent applications and prospects. *npj Computational Materials*, *3*(1), 1-13. doi:10.1038/s41524-017-0056-5

*Rodríguez-Martín, M., Rodríguez-Gonzálvez, P., Sánchez-Patrocinio, A., & Sánchez, J. R. (2019). Short CFD simulation activities in the context of fluid-mechanical learning in a multidisciplinary student body. *Applied Sciences*, *9*(22), 4809. doi:10.3390/app9224809

*Rodriguez-Muniz, L. J., Bernardo, A. B., Esteban, M., & Diaz, I. (2019). Dropout and transfer paths: What are the risky profiles when analyzing university persistence with machine learning techniques? *PLoS One, 14*(6), e0218796. doi:10.1371/journal.pone.0218796

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, *26*(2), 582–599.

Schunk, D. H. (2020). *Learning theories an educational perspective* (8th ed.) Boston, MA: Pearson.

Shi, B., & Weninger, T. (2017). ProjE: Embedding projection for knowledge graph completion. *Proceedings of the Thirty-first AAAI Conference on Artificial Intelligence, 31*(1), 1236–1242.

Sonderlund, A. L., Hughes, E., and Smith, J. (2019). The Efficacy of learning analytics interventions in higher education: A Systematic review. *British Journal of Educational Technology*, 50(5), 2594–2618.

*Spatiotis, N., Perikos, I., Mporas, I., & Paraskevas, M. (2020). Sentiment analysis of teachers using social information in educational platform environments. *International Journal on Artificial Intelligence Tools*, *29*(02), 2040004. doi:10.1142/S0218213020400047

*Sultana, S., Khan, S., & Abbas, M. A. (2017). Predicting performance of electrical engineering students using cognitive and non-cognitive features for identification of potential dropouts. *International Journal of Electrical Engineering & Education, 54*(2), 105-118.

The White House, Office of the Press Secretary. (2015). *FACT SHEET: President Obama's Precision Medicine Initiative*. Retrieved from https://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative

Tseng, J. C., Chu, H. C., Hwang, G. J., & Tsai, C. C. (2008). Development of an adaptive learning system with two sources of personalization information. *Computers & Education*, *51*(2), 776-786.

*Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An Overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & Education, 143,* 103676. doi:10.1016/j.compedu.2019.103676

*Tsai, S.-C., Chen, C.-H., Shiao, Y.-T., Ciou, J.-S., & Wu, T.-N. (2020). Precision education with statistical learning and deep learning: A Case study in Taiwan. *International Journal of Educational Technology in Higher Education, 17*(1), 1-13.

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018) The Current landscape of learning analytics in higher education. *Computers in Human Behavior, 89,* 98–110.

*Wakelam, E., Jefferies, A., Davey, N., & Sun, Y. (2019). The Potential for student performance prediction in small cohorts with minimal available attributes. *British Journal of Educational Technology, 51*(2), 347-370.

*Wang, L., Hu, G., & Zhou, T. (2018). Semantic analysis of learners' emotional tendencies on online MOOC education. *Sustainability, 10*(6), 1921. doi:10.3390/su10061921

Williamson, B. (2019). Digital policy sociology: Software and science in data-intensive precision education. *Critical Studies in Education*, 1-17. doi:10.1080/17508487.2019.1691030

Wu, D., Jennings, C., Terpenny, J., Gao, R. X., & Kumara, S. (2017). A Comparative study on machine learning algorithms for smart manufacturing: Tool wear prediction using random forests. *Journal of Manufacturing Science and Engineering*, *139*(7), 071018. doi:10.1115/1.4036350

*Wu, J. Y., Hsiao, Y. C., & Nian, M. W. (2020). Using supervised machine learning on large-scale online forums to classify course-related Facebook messages in predicting learning achievement within the personal learning environment. *Interactive Learning Environments*, *28*(1), 65-80.

*Xiao, M., & Yi, H. (2020). Building an efficient artificial intelligence model for personalized training in colleges and universities. *Computer Applications in Engineering Education*, 1–9. doi:10.1002/cae.22235

Xie, H., Chu, H. C., Hwang, G. J., & Wang, C. C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: A Systematic review of journal publications from 2007 to 2017. *Computers & Education*, *140*, 103599. doi:10.1016/j.compedu.2019.103599

*Xing, W., & Du, D. (2018). Dropout prediction in MOOCs: Using deep learning for personalized intervention. *Journal of Educational Computing Research, 57*(3), 547-570.

*Xing, W., Popov, V., Zhu, G., Horwitz, P., & McIntyre, C. (2019). The Effects of transformative and non-transformative discourse on individual performance in collaborative-inquiry learning. *Computers in Human Behavior*, *98*, 267-276.

*Xu, J., Moon, K. H., & van der Schaar, M. (2017). A Machine learning approach for tracking and predicting student performance in degree programs. *IEEE Journal of Selected Topics in Signal Processing, 11*(5), 742-753.

U.S. Department of Education, Office of Educational Technology (2017). *Reimagining the role of technology in education: 2017 national education technology plan update.* Retrieved from https://tech.ed.gov/files/2017/01/NETP17.pdf

Voyant, C., Notton, G., Kalogirou, S., Nivet, M. L., Paoli, C., Motte, F., & Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A Review. *Renewable Energy*, *105*, 569-582.

Yang, S. J. H. (2019). Precision education: New challenges for AI in education [conference keynote]. In *Proceedings of the 27th International Conference on Computers in Education (ICCE)* (pp. XXVII-XXVIII). Kenting, Taiwan: Asia-Pacific Society for Computers in Education (APSCE).

Zawacki-Richter, O., Marín, V. I., Bond, M., and Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), 1-27. doi:10.1186/s41239-019-0171-0

Zhai, X., Yin, Y., Pellegrino, J. W., Haudek, K. C., & Shi, L. (2020). Applying machine learning in science assessment: A Systematic review. *Studies in Science Education*, *56*(1), 111-151.

*Zhou, Y., Huang, C., Hu, Q., Zhu, J., & Tang, Y. (2018). Personalized learning full-path recommendation model based on LSTM neural networks. *Information Sciences*, *444*, 135-152.

Wu, J.-Y., Yang, C.C.Y., Liao, C.-H., & Nian, M.-W. (2021). Analytics 2.0 for Precision Education: An Integrative Theoretical Framework of the Human and Machine Symbiotic Learning. *Educational Technology & Society*, *24* (1), 267-279.

Analytics 2.0 for Precision Education: An Integrative Theoretical Framework of the Human and Machine Symbiotic Learning

Jiun-Yu Wu[1*], Christopher C.Y. Yang[2], Chen-Hsuan Liao[1] and Mei-Wen Nian[1]
[1]Institute of Education, National Chiao Tung University, Taiwan, R.O.C. // [2]Graduate School of Informatics, Kyoto University, Japan // jiunyu.rms@gmail.com // yang.yuan.57e@st.kyoto-u.ac.jp // abugu103@gmail.com // anke801211@gmail.com
[*]Corresponding author

ABSTRACT: This methodological-theoretical synergy provides an integrative framework of learning analytics through the development of the human-and-machine symbiotic reinforcement learning. The framework intends to address the challenges of the current learning analytics model, including a lack of internal validity, generalizability, immediacy, transferability, and interpretability for precision education. The proposed framework consists of a master component (the brain) and its four subsuming components: social networking, the smart classroom, the intelligent agent, and the dashboard. The brain component takes in and analyzes multimodal streams of student data from the other components with the model-based reinforcement learning, which forms policies of adequate actions that maximize the long-term rewards for both the human and machine in the seamless learning environment. An example case plan in advanced statistics was demonstrated to illustrate the course description, data collected in each component, and how the components meet different features of the smart learning environment to deliver precision education. An empirical demonstration was provided using some selected mulitmodal data to inform the effectiveness of the proposed framework. The human-and-machine symbiotic reinforcement learning has theoretical and practical implications for the next-generation learning analytics models and research.

Keywords: Reinforcement learning, Learning analytics, Symbiotic learning, Smart learning environment, Precision education

1. Introduction

The advent of Information and Communication Technologies (ICT) has yielded explosive and drastic changes in human thinking and decision making (Jordan & Mitchell, 2015). Among the technologies, Machine Learning (ML) is the core of artificial intelligence and data science; via supervised, unsupervised, or reinforcement learning, machines imitate human behavior and thinking and excel in an automatic recursive process (Jordan & Mitchell, 2015). ML's affordance offers us insights into the advancement and development of the next generation of Learning Analytics (LA) to implement precision education (Wu et al., 2020). Mainly, precision education is "an approach to research and practice that is concerned with tailoring preventive and intervention practices to individuals based on the best available evidence" (p. 4, Cook et al., 2018). Whereas, LA is defined as using data from educational institutes to construct prediction models for improving the student learning process (Wu, 2020). More specifically, LA measures, gathers, analyzes and returns students data and information content in order for stakeholders of education to better investigate and understand the environment within and outside the learning entity where learning takes place for personalized feedback and instruction (Siemens & Baker, 2012). Thus, harnessing the power of LA, precision education may be more promising and ready to benefit students and instructor in their learning and teaching with engaging, flexible, adaptive, and personalized diagnosis and intervention. Nevertheless, new challenges arise as to how artificial intelligence and ML can be well applied in LA to achieve precision education for capturing the heterogeneity among learners and facilitating students' learning performance and instructors' teaching quality (Yang, 2019).

Notably, LA has been one of the most innovative areas to accomplish precision education with its leverage in the construction of prediction model and student learning database as well as the analysis of a vast amount of qualitative and quantitative multimodal data for personalized learning (Blikstein & Worsley, 2016). Despite the advantages, there are deficiencies in the current LA that may limit its application for precision education. These deficiencies include lacking in internal validity, immediacy/automated feedback, and generalizability. Specifically, the above-claimed effects of learning and teaching based on learning analytics are post-hoc in nature. The progress and application of the analytical results are not synchronous. Namely, the models or indicators obtained from the previous students can only be used or applied to the students in the next stage or generation, whereby lacking internal validity and failing to inform instructors of their students' recent status.

Meanwhile, the result of learning analytics cannot be used for immediate and real-time diagnosis and automated feedback at the current stage, whereby lacking immediacy (Sedrakyan et al., 2018). Specifically, the criterion-based validity check, usually using the students' grades, limits the possibility to use all the available indicators and all possible actions exhaustively. As a result, it may hinder external validity and damage the generalizability of the LA model.

Besides the constraints mentioned above, additional challenges for the current LA research include the transferability between the different learning systems and translation of the LA results to the instructor or practitioner's language (Baker, 2019). Moreover, researchers have noted that learners' lack of access to their learning data reduces their opportunities to make sense of their learning process and hinders their metacognition and self-regulation (Kitto et al., 2017; Wu, 2014; Wu & Peng, 2017). The limitations and constraints of the current learning analytics models mentioned above demand the attention of the LA research community for achieving precision education.

This study intends to provide an integrative theoretical framework of learning analytics for precision education through the human-and-machine symbiotic reinforcement learning (RL). We contend that the RL can be the core application of the next-generation LA, namely LA2.0, to address the constraints of internal validity, generalizability, immediacy, transferability, and interpretability for the current state of learning analytics so that LA2.0 can be readily utilized to enhance precision education. Thus, we aim to answer two research questions:

- What are the possible components in the integrative theoretical framework of learning analytics for precision education?
- What is the efficiency of this proposed learning analytics framework in modeling learning performance for precision education?

Below we provide the theoretical underpinning of the framework, including smart learning environment, learning analytics and reinforcement learning, and learning analytics 2.0: the framework for precision education.

2. Literature review

2.1. Building a smart learning environment for precision education based on the affordance of adaptive technologies

Smart Learning Environment (SLE) is built upon adaptive technologies that satisfy learners of different backgrounds and engage them in context-aware learning activities that suit their goals (Spector, 2014). In higher education, social media is one of the applications of adaptive technologies that enable learners to create their unique Personal Learning Environment (PLE, Dabbagh & Kitsantas, 2012; Wu, 2017). Premised on social media, the PLE encompasses Learning Management Systems (LMS, e.g., Moodle, Canvas, or Blackboard) and thus is more open and flexible than the close systems. Learners can create their PLE using various applications: blogs, wikis, google apps/calendars, dropbox, YouTube/Flickr, etc. to achieve their learning goals considering their interests, preferences, emotions, and attitudes. Learners can also share, communicate, and collaborate by extending their PLE to form a personal learning network or community with experts and more knowledgeable peers. Despite PLE's advantages, people's limited attention resources are significantly challenged by fun and exciting events and activities, such as friends' tweets, photos, fun games, and videos (Wu, Online first; Wu & Xie, 2018). Thus, learning premised on social media may be a double-edged sword due to students' distracted attention and poor self-regulatory strategies (Wu, 2015, 2017; Wu & Cheng, 2019). These external threats in the PLE create an urgency to build an SLE with analytical evidence for students' autonomous and self-directed learning. An SLE should be able to engage students in learning with its technological affordance and meanwhile facilitate students to plan and monitor their learning progress actively.

The extent of an SLE can be categorized by its necessary, strongly desired, and likely features (Spector, 2014). Necessary features imply that the SLE should have evidence to support its effectiveness and efficiency for students' autonomous learning based on diverse and large-scale samples. In terms of the highly desired, engaging, flexible, adaptive, and personalized are the four main features. The SLE may be engaging in arousing and maintaining students' motivation, attention and engagement. Meanwhile, it can be flexible to accommodate changes in the course (e.g., adding new members, changes in learning goals) and being adaptive to students' abilities, interests, or cognitive styles to provide personalized instruction and feedback for those falling behind or progressing ahead. The necessary and highly desired features can benefit from the affordances of Web 2.0 and Web 3.0. Students can actively create their responses and construct knowledge schema with their peers and the instructor synchronously or asynchronously, with the teacher orchestrating and coordinating students' collaboration (Gerstein, 2014). Based on connectivism theorized by Siemens (2005), teaching and learning may represent a combination of numerous activities and networked and interwoven complex relationships. When students share information or post/reply questions and comments by interacting with peers or instructors, they connect their current state of knowledge to form knowledge between ideas, concepts, and domains via specialized nodes and information sources (Wu & Nian, 2021). Thus, built on connectivism (Siemens, 2005), learners of Web 3.0 applications can be creators of content knowledge. They can share their intellectual artifacts with others in networked learning and connect resources, persons, communities, and applications/tools relevant to their learning via the far-reaching web (Berners-Lee et al., 2001). Under this scenario, learning is highly autonomous and self-determined, with teachers taking the role of a coach or cheerleader (Gerstein, 2014).

As for Web 4.0., researchers suggest a symbiotic web where humans and machines have an interdependent and coexisting relationship (Gamberini & Spagnolli, 2016). The concept of human-machine symbiotic relations offers immense possibilities for human learning. In light of Spector's likely features for an SLE: conversational, reflective, innovative, and self-organizing, we envision that the SLE built upon the affordances of Web 4.0 can engage learners in dialogs for problem-solving, create learning progress reports for students' evaluation of their performance, use technologies in innovative ways to support students' learning, and help improve students' performance over time by automatically managing resources and collecting/analyzing data from the learning ecology. To implement these likely features in the SLE requires comprehensive and penetrating learning analytics encompassing all possible aspects of data about students' learning. Below we discuss the now and future of learning analytics regarding its role in the SLE.

2.2. Learning analytics and reinforcement learning

LA's ultimate purpose is to provide precision education for individualized instruction and feedback to enhance students' motivation and achievement with student-related data gathered from multiple sources (Romero & Ventura, 2020; Siemens & Baker, 2012). Nevertheless, the current LA models may constrain their use for such purposes. Specifically, the current LA studies are mostly using off-line training data for prediction based on static models (Nishihara et al., 2017). They cannot provide real-time and immediate feedback or support to facilitate student learning due to programming flexibility and performance limitations. Reinforcement Learning (RL), a broader paradigm of machine learning, may address the constraints mentioned above. RL can fuse and react to multiple sensory data from various input streams, conduct micro-simulations continuously, and figure out the next step (Nishihara et al., 2017). Compared with other machine learning algorithms, RL focuses more on goal-directed learning via interacting with the environment (Sutton & Barto, 2018). The robot has clear goals, perceives the environment, and selects an action to respond to or change the environment. RL differs from supervised or unsupervised learning because it learns the behavior based on the feedback obtained from iteratively interacting with the environment. Thus, RL resembles the conditioning learning of humans or animals (Sutton & Barto, 2018) to master the skill. The RL-based agent/robot applies the concept of "reinforcement" in behaviorism, where humans make decisions based on the state of the current environment and select the corresponding action. Once the environment rewards them, they may maintain and adjust their policy to maximize their long-term reward.

There are four elements in the RL: policy, reward signal, value function, and model (Sutton & Barto, 2018). The policy perpetuates the RL robot's action with the standard strategy to maximize the value function. The reward signal is the value obtained by the RL robot from the environment to evaluate its action's performance. The value function is the RL robot's expected values/rewards across all the possible actions; the RL robot obtains its value function by continually updating with the latest parameters. Before the RL robot executes its action, models could help predict what reward the environment may give to decide its strategy use. A model-free RL robot can be used for explicit trial-and-error searches (van Otterlo & Wiering, 2012). Alternatively, mode-based RL agents can be applied to reflect the action that gains more weights in reward from the environment (Hester & Stone, 2012). This study proposed the human-machine symbiotic learning analytics framework based on the mode-based RL algorithm as an updated version of the current learning analytics model. The framework may be able to bolster the likely features in the SLE for conversational, reflective, and self-organizing and innovative learning via its low latency and high throughput to support online simulations and the streaming sensory input (Nishihara et al., 2017).

3. Learning analytics 2.0: The framework for precision education

This study aims to address the problems and constraints of the current Learning Analytics to construct a human-machine symbiotic reinforcement learning framework for precision education. A summary of aspects that entail the human-machine symbiotic relationship is exhibited in Table 1. Specifically, humans can provide rules to supplement the building of domain knowledge and just-in-time information to ease the computational complexity with more efficient and effective training results (Sutton & Barto, 2018). Humans can also collaborate with the RL robot by providing adversarial training for valid value evaluations to gain maximum rewards despite extensive simulation training (Pinto et al., 2017). In terms of the distinct style of knowledge building, the result of machine learning training is complex and is rendered powerless unless it can be interpreted by human experts (Vellido et al., 2012). Thus, the learning analytics results with experts' input can be displayed via interactive visualization and dashboards to provide learners and instructors with meaningful interpretation (Aljohani et al., 2019). Moreover, as experts, humans can provide domain knowledge regarding RL's susceptibility to partial information (Abbeel & Ng, 2004).

Table 1. The symbiotic relationship between the machine (the RL algorithm) and human

Aspects that entail the symbiotic relationship	What RL can do	Constraints of RL	How humans can collaborate with RL
The domain of knowledge building	Needs no instruction/knowledge.	Rules needed	Provide rules
Speed of computation	Trial-and-error achieved in seconds.	Not a fast learner without supercomputers due to computational complexity	Provide just-in-time information to help the RL learn faster
The capability of knowledge building	AIs with RL generate their knowledge.	Need valid value evaluation	Provide adversarial training for valid value evaluation
Distinct styles of knowledge building	Has its style of knowledge building.	Difficult to be explained	Interactive visualization and dashboard display based on suggested actions and estimated rewards
Susceptibility to partial information	Could be generalized to other situations.	Not effective with only partial information	Serve as an expert or a community of experts to provide domain knowledge.

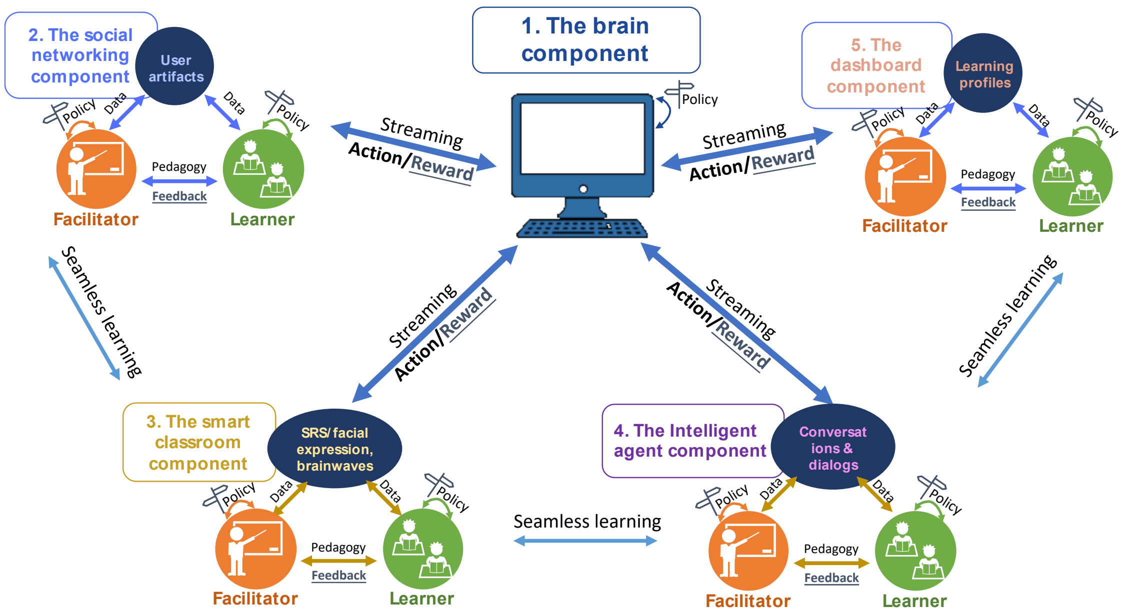


Figure 1. The integrative framework of the human-machine symbiotic learning within the smart learning environment

Harnessing the power of RL, we propose an integrative LA framework for precision education, consisting of five components. We illustrate the association of the five components as shown in Figure 1 and elaborate on the functionality of the components in the following subsections.

3.1. The brain component

The brain is the master component based on connectivism (Siemens, 2005) and RL (Sutton & Barto, 2018). With analytical techniques (e.g., statistics, data mining, supervised/non-supervised machine learning techniques), the brain takes in multimodal streams of students' data in the hybrid seamless instructional settings to build the human-machine symbiotic learning model. The information streams may include automated coding of supervised data and model-free data and metadata (e.g., the number of messages and comments posted on the wall) from the social network component, class attendance, facial expression, physical movement, physiological signals, and instant assessment responses from the smart-classroom instruction component, conversations, and dialogs from the intelligent agent component, and human-computer interactions and information-exchanging visualizations from the dashboard component (e.g., students set goals and regulate their learning based on dashboard feedbacks).

The brain establishes policies to select an action in response to the learning environment with strategies that maximize the value function for the greatest reward; the reward would then inform the brain of the result of the action in a recursive manner for self-organizing. The brain component is designed with 1) low latency and high throughput, 2) dynamic task creation, heterogeneous tasks, and arbitrary dataflow dependencies, and 3) transparent fault tolerance and debuggability and profiling to meet the performance, execution, and practical requirement of emerging real-time machine learning (Nishihara et al., 2017).

3.2. The social networking component

The second component is the social networking component characterized by students' cognitive, affective, and social artifacts created on social media (Lee & Wu, 2013). Specifically, the context of social media use or what and how students are using social media plays a significant role in students' outcomes. Researchers exhibited that students with messages and posts endorsed by two or more robots as statistics-relevant had higher final course grades; moreover, students failing the course had significantly fewer messages endorsed by three robots as statistics-relevant than those who passed (Wu et al., 2020). Therefore, content or the interaction context is the core element that may predict students' performance in social networked learning. Moreover, social presence or the ability to perceive others in the online learning environment is also associated with learners' satisfaction and perceived learning (Dabbagh & Kitsantas, 2012). Thus, learners' posts and comments among peers and instructors and their reactions to others' messages (i.e., emojis, or the buttons of Like, Haha, Love, Wow, Sad or Angry) are essential cognitive and affective artifacts in social networked learning.

By using both supervised text classification and RL enabled student data streaming, the social network component has the potential to classify learners' messages based on the cognitive level or sentiment analysis (Cambria et al., 2017) and send those data along with the metadata (e.g., Facebook reactions) on the social media in real-time to the brain component in order to construct policies in response to the dynamic environment for maximum reward in learning.

3.3. The smart classroom component

The third component is the smart classroom component. It focuses on the observation, monitoring, and interaction among peers and instructors and pedagogical adjustment in the face-to-face classroom using multimodal data for learning analytics. Grounded in connectivism (Siemens, 2005), learning resides in the exchanges of diverse opinions. In the classroom setting, learners can learn from reciprocal inquiry and dialectic learning with their peers and instructors. However, in higher education, the class size causes a barrier in the teacher-student interaction, with negative correlates such as higher dropout rate and retention (Bettinger & Long, 2018). Student response systems (SRS) can improve students' learning, motivation, and engagement and bring more opportunities to improve discussion and interaction between students and teachers and among students for educational purposes (Wu et al., 2019). By implementing formative assessments via SRS adaptively according to the analytical result streamed back from the brain component, the instructors can identify misconceptions in students learning and provide immediate scaffolds or direct explanations on the idea. Besides, students' opinions can be more openly expressed via SRS against conformity or shyness.

In addition to students' responses to the formative assessment via SRS, their facial expressions, physical movements, and physiological signals (e.g., skin conductance and brainwaves) may also be recognized to detect their learning emotions. Advances in machine learning have made real-time facial expression recognition of emotions feasible. Using automated recognition of facial recognition, researchers found that students' upper face movements are related to engagement, frustration, and learning, while mouth dimpling positively predicts learning and self-reported performance (Grafsgaard et al., 2013). Real-time facial and head gesture recognition can also successfully identify students sleeping, yawning, and smiling as well as nodding, shaking, and tilting with high accuracy (Deshmukh et al., 2018). Besides, in face biometric systems, the non-contact algorithms have been carried out for virtually physiological signal detection, e.g., pulse rate registration, directly from face images captured from motion videos (Lewandowska et al., 2011). Moreover, students' physiological signals such as skin conductance also exhibited a positive correlation with their self-reported mental efforts in solving ill-structured problems (Larmuseau et al., 2019). These non-verbal signals can provide the instructor with valuable information about students' attentional and cognitive states, engagement, and motivation for the instructor's pedagogical adjustment to facilitate students' learning and help manage student attendance (Kar et al., 2012).

Students' responses via SRS, facial, physical movements, and physiological signals can also be coded, classified, and preprocessed by edge computing mobiles and terminals (Shi et al., 2016). The coded responses can be sent to the brain component via the streaming technology to establish, maintain, or adjust policy for obtaining the maximum reward.

3.4. The intelligent agent component

The fourth component is the intelligent agent component. Though the instructor and peers can interact with the learner and exchange ideas and opinions, the time and space constraints would limit the currency of knowledge creation and co-construction. To address the possible limitation of failing to provide just-in-time feedback and scaffolds, we propose the intelligent agent component that can work as a bridge in the relationship among the content, peer, and the instructor for knowledge transfer and creation. Intelligent agents are applications of artificial intelligence built upon machine learning and natural language processing for personal, conversational, and engaging ways of learning. Intelligent agents can serve as the role of the more knowledgeable other in the zone of proximal development (Vygotsky, 1987) to assist in completing the task, answering conceptual questions, or prompting learners' reflection or metacognition in a conversational way. Nonetheless, most intelligent agents are still based on fixed rules and may not provide feedbacks given students' characteristics and needs (e.g., Pereira, 2016).

RL and task-based design empower the proposed intelligent agent component. It uses streaming student data from the brain component, the social networking component, and the smart classroom component to provide real-time scaffolding on students' misconceptions and guide students' critical thinking to achieve their learning goal. The conversation, dialogs, and interaction between the learners and intelligent agents can also be exported to the brain component to improve the human-machine symbiotic reinforcement learning mechanism.

3.5. The dashboard component

The fifth component is the dashboard component, which is in charge of multiple sensory data gathering, cleaning, storage, and management and visualizing each student's learning profile. Research has documented the advantages of the learning analytic dashboards, such as to identify students' role in online learning or their interaction with others (Ferguson & Shum, 2012), to enhance the adviser-student dialog via visualizing the study progress and comparison with peers for discussion and argumentation (Charleer et al., 2018), and to support students' self-regulated learning with corresponding features of self-monitoring and self-assessment complemented with customized feedbacks (Schumacher & Ifenthaler, 2018). Students can also compare their class participation and performance with the class's average performance or the best-achieving students by available factors (Aljohani et al., 2019). Linking the learning analytics dashboard display with learning science concepts, researchers propose that the dashboard design should help students in the planning, performance, and adaption phases of their learning (Sedrakyan et al., 2018). Nevertheless, there are still challenges for the learning analytics dashboard design, including 1) difficulty in modeling the dynamics of learning, 2) failure to taking into account of learner characteristics, 3) limitation in the modality of student data, and 4) inclusion of student data from a single platform (Sedrakyan et al., 2018).

The current study proposes a human-machine symbiotic reinforcement learning framework using the fine-grained and accumulated learning progress data to investigate and predict students' next-stage action based on their previous moves. Therefore, it may help resolve the problem of lacking internal validity and avoid providing feedback based on the previous sample's data to the students of the ongoing section.

4. Empirical demonstration and discussion

This section provided an example course plan in a graduate-level advanced statistics course to depict the integrative theoretical framework. As a demonstration for the proposed integrative LA framework, we then conducted an empirical analysis with available data from the four subsuming components obtained from an advanced statistics course. Though the analytical model reported here was still post-hoc in nature, we can envision the development of a human-machine symbiotic learning when all the multimodal multiple-source data are incorporated and streamed to the brain component. The brain, which is powered by the RL mechanism and statistical modeling, can then establish real-time policy and decision making for precision education.

4.1. Description of the components in the example case plan

In response to our first research question, we described the course component and data that can be collected in each component to demonstrate how each component meets different features of a smart learning environment for precision education in Table 2. The brain's primary function is to establish policies with strategies that maximize value functions for the greatest reward and meet the performance, execution, and practical requirement of reinforcement learning (Nishihara et al., 2017) so that it can coordinate all the components in the integrative framework. Specifically, the advanced statistics instructor can design the course using a flipped learning approach; therefore, in the social networking component, students can preview the course videos on Youtube or LMS before the class. Students can also review the video lectures afterward based on their study pace. The instructor creates a statistics learning Facebook group to share information and seek help via posts and comments for seamless learning and self-regulation.

Table 2. The sample case of an integrative framework of the human-machine symbiotic learning within the smart learning environment

The Master component	The Brain Component			
	• Built upon the smart learning environment (Spector, 2014), connectivism (Siemens, 2005) and the RL algorithm (Sutton & Barto, 2018) • Receive multimodal streams of students' data • Develop the human-machine symbiotic learning model			
Subsuming Components	The social networking component	The smart classroom component	The intelligent agent component	The dashboard component
Course component description	• Before/after the class: The instructor provides course videos for previewing. Students can review the video if necessary. • Before/during/after the class: Students are encouraged to share information and seek help via the Facebook group for seamless learning and self-regulation.	• During the class: The instructor integrates the student response system (e.g., Kahoot!) for quizzes to monitor students' comprehension of the statistics concepts. • Students' facial expressions, physical movement, and physiological signals can be recorded and	The intelligent agent is a students' personalized tutor that can address each student's specific needs or statistics misconceptions based on his/her familiarity with the course materials, e.g., performance on the student response system, posts/comments	The dashboard displays students' statistics learning progress across different platforms (e.g., amount of video viewing, discussion participation, interaction among peer, weekly quiz performance). It takes information from all the previous components plus additional psychological assessments to tailor each student's learning program. Students are allowed to set their learning goals.

		transmitted to the brain component for instant recognition of students' attention and emotional status toward a specific concept of statistics learning.	on the Facebook group, and classroom engagement.	Automated and individualized feedbacks are given considering all available student pre-existing conditions.
Data collected	Course video viewing time on LMS and posts, comments, and reactions (emojis) on the Facebook group	Students' instant responses to the test items, facial expression, physical movement, physiological signals	All available data from the social networking component and the smart classroom component	All available data from the social networking component, the smart classroom component, and the intelligent tutor component plus psychological assessments of students' learning preferences, strategies, and habits, etc.
Features of smart learning environment	Necessary features: scalable, effective, efficient, and autonomous.	Highly desired features: engaging, flexible, adaptive, and personalized	Likely: conversational, reflective, innovative, and self-organizing	Likely: conversational, reflective, innovative, and self-organizing

In the smart classroom component, the instructor can create a quiz bank using the gamified student response system (e.g., Kahoot!) and adaptively use the quiz items based on the brain component's action to monitor students' comprehension of the statistics concepts. Students' facial expression (e.g., yawn, frown), physical movement (e.g., stretch, sleep), physiological signals (e.g., pulse, skin conductance, brainwaves) can be recorded. The signals can then be transmitted to the brain component for instant recognition or automated classification of students' attention and emotion status (e.g., frustrated, confused, happy, bored, or stressful) toward a specific concept of statistics learning. The use of the student response system was shown to increase student engagement in the class. Moreover, instant automated classification of student-related signals can allow the instructor to provide flexible, adaptive, and personalized feedback for each student and ease the instructor's cognitive load given a large class size (Stowell et al., 2010).

The intelligent agent component is designed based on rules and all available information from the social networking component and the smart classroom component. Thus, students' personalized tutor can address each student's specific needs or statistics misconceptions based on his/her unique records. The personal records consist of students' familiarity with the course materials (course video viewing time), performance on the student response system, posts/comments on the Facebook group, and classroom engagement. For example, a student may get a quiz item wrong in the class, or ask for clarification about statistics concept on the Facebook group. Then, the intelligent agent would notice the student's possible weaknesses and provide relevant materials/concepts/questions of different difficulty levels for the student to think or work on and scaffold the learning process. The intelligent agent is an innovative technology that can form a conversational and reflective way of learning and is capable of self-organizing the learning materials and sequences.

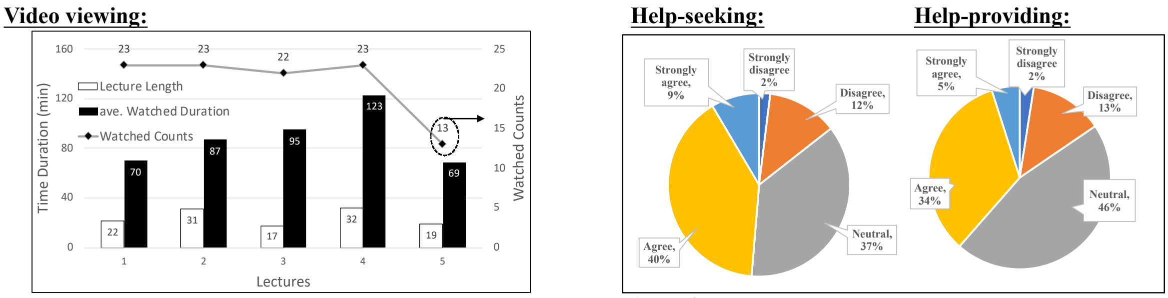
The dashboard component displays students' statistics learning progress across different platforms (e.g., amount of video viewing, discussion participation, interaction among peer, weekly quiz performance). It takes information from all the previous components plus additional psychological assessments (e.g., Internet use habits and epistemic beliefs (Lee, 2018, 2021)) to tailor each student's learning program better. Students are allowed to set their learning goals and regulate their learning using the dashboard components (Sedrakyan et al., 2018). Automated and individualized feedbacks for goal-setting or self-regulation are given considering all available student pre-existing conditions. The feedback distribution algorithm is based on the RL policy that yields the most considerable reward and can avoid students' unrealistic goal-setting or self-evaluation (too high or low) and help them self-regulate their learning. The student learning profiles provided by the dashboard can also be used by the instructor to adjust their course plan and teaching strategies.

4.2. The empirical demonstration of an LA based on an actual course

In response to our second research question, we simulated a prediction model in the brain component on students' learning performance to test the efficiency of the LA model. Specifically, we modelled the prediction effectiveness of an LA based on some selected multimodal and multiple-source data from an actual course. Participants were 23 graduate students taking an advanced statistics course. In the social networking component, the duration of each individual's time spent on each part of a video lecture was recorded, including the time paused and replayed. Students also provided ratings of their perceived help-seeking desire and help-providing capability, as well as a brief reflection of what they have learned and they wanted to share after viewing the lecture videos. Moreover, students' posts and comments on the Facebook group were classified into statistics relevant or irrelevant using supervised machine learning (Wu et al., 2020). In the future, we hope to replace the classification process by the RL with expert's rule to establish real-time policies and provide immediate personalized feedback. Further, in the smart learning classroom, students' responses to the in-class quizzes was recorded using the gamified instant response system (i.e., Kahoot!). We created a proxy variable for the intelligent agent component by dividing the number of responses a student received from the instructional team by the total messages the students posted. The proxy variable may function similarly as the adaptive tutoring when a student signals the need for help. The proxy variable was created in place of the personalized feedback in the intelligent agent component for demonstration purposes; however, it only considered feedback received from the instructor and teaching assistants. Thus, more research is needed to advance the development of intelligent agents based on rules and multimodal and multiple-source inputs.

As a visualization to help students' self-regulation, we built a dashboard based on students' video viewing and perception/reflection data. As illustrated in Figure 2, the dashboard presented the duration of each lecture, the average students' time spent viewing each lecture, and the number of students who watched the video on the upper left-hand side. On the upper right-hand side, the pie charts presented the percentages of students in terms of the degree of their perceived help-seeking desire and help-providing capability. The bottom left-hand side and bottom-right hand sides exhibited all students' qualitative accounts of their brief learning reflection and perception about the lectures, respectively. The visualizations can be provided to students on a weekly basis; thus, the dashboards may enhance students' understanding of their learning in comparison to the whole class for goal-setting and self-regulation. Instructors can also use the learning analytics dashboards to monitor each student' learning progress over time and identify students who are in need of help at the earliest stage to prevent them from falling behind or failing the course.

The Learning Record of a Unit: Multiple Regression



I have learned:

- There's even a space before "Orc". If TA hadn't mentioned that, I would have to waste a lot of time for debugging my code.
- The ultimate purpose of regression is the effective prediction that does not need to wait for the results to be read out.
- Multiple Regression is different from simple linear regression. More independent variables are added in while the dependent variable remains to be only one.
- Regression + residual = real data. We should focus on the standardized regression coefficient. Under the consideration of theory, it's better to put more than only one meaningful independent variables into the regression model. To some extent, it represents a more realistic model.
- The larger the tolerance, the better the model. On the other hand, the smaller the VIF, the better the model.
- $SS_{total} = SS_{regression} + Ss_{error}$. We should refer the standardized coefficients.

Anything want to share:

- It's interesting~~
- (Blank)
- I'm still confused about the post-hoc after the linear regression analysis as shown on page.12 in the handout.
- It's so great that I can adjust the video to a faster speed. When it comes to the film quality, however, it only provides 480p, which is a bit obscure.
- (Blank)
- Not yet, I'm still digesting the knowledge.
- I'm in favor of this topic. I believe it's very useful.
- I have to rush to the next video. Bye~
- I still can't understand the concept about the collinearity. Hope there will be more explanation in the future.
- Good job. Very clearly.

Figure 2. The illustration of a learning analytics dashboard

To simulate the analytical functioning of the brain component, we constructed a regression model using the multimodal and multiple-source data described above in predicting students' final course grades, controlling for students' gender and prior knowledge in statistics. The regression result was shown in Table 3. Students' prior knowledge, in-class quizzes (Kahoot!), the number of messages classified by the ML as relevant, and their perceived help-seeking desire were associated with higher final course grades with statistical significance. The model explained 80.83% of the variance (adjusted R2) in the scores of the final course grades.

The empirical demonstration provided an initial exploration of the multimodal data from different components of the LA framework in predicting learning achievement. To yield internal validity and immediacy in personalized feedback in the future, the multimodal data can be collected on a weekly basis (or a more fine-grained time period) to form and adjust the RL policies based on students' subsequent formative assessments (e.g., in-class quizzes, midterm, or homework) and expert feedback. The policies can then be used for classification or association of students' learning progress for decision making, such as signaling of wheel-spinning or at-risk of failing.

The empirical demonstration based on an actual course had implications for precision education. First, learners' prior knowledge explained a significant portion of their final course grades. Understanding learners' initial level of the content knowledge can inform the instructor of students who are more likely to need assistance. Next, the Kahoot quizzes and ML classified messages provided additional clues regarding students' learning progress as the semester course went. Learners' with poorer quiz scores and fewer messages endorsed by the ML algorithm may flag at risk of academic failure. Finally, learners' help-seeking tendency may represent their desire to close their knowledge gap. Whereas, those with low help-seeking tendency may signal a reactive learning attitude, and thus, may require more instructor's attention and assistance. Suggestions can then be made to inform instructors of personalized intervention and enhance students' understanding of their learning for self-regulation and goal-setting in an automatic process. The analytical findings are promising and warrant the advancement of LA2.0 for precision education.

Table 3. The multiple regression analysis with multimodal inputs on students' final course grades

	B	*SE*	*β*	*t*	*p*	*VIF*
(Constant)	3.773	12.052		0.313	.758	
Male	2.765	1.904	0.138	1.452	.164	1.171
Prior Knowledge	0.234	0.088	0.310	2.644	.016*	1.784
Kahoot!	0.565	0.127	0.634	4.469	<.001**	2.615
ML_classified Message	0.375	0.135	0.279	2.775	.012*	1.313
Proxy_intelligent agent	-0.025	0.064	-0.038	-0.392	.699	1.204
Help seeking	3.912	1.805	0.255	2.167	.044*	1.792
Help providing	2.828	1.488	0.181	1.901	.073	1.175

Note. $^{*}p < .05$; $^{**}p < .01$.

5. Conclusion

Learning Analytics is one of the emerging learning technologies that apply learner-generated data and all other related information to provide personalized instruction and help learners adapt their learning in the technology-enhanced environment. The proposed human-and-machine symbiotic reinforcement learning has both theoretical and practical implications for the next-generation learning analytics models to implement precision education. From a theoretical perspective, the integrative Learning Analytics framework premised on the RL mechanism with human expert input (i.e., model-based RL) addresses the RL algorithm's limitations by providing rules and just-in-time human expert knowledge exchange. Therefore, RL robots can learn faster without the need for domain knowledge and computation with supercomputers. Valid value evaluations meaningful to humans can also be achieved by providing adversarial training for the RL robot to take appropriate policies of a series of actions to gain the maximum reward (Pinto et al., 2017). As experts, humans can also provide domain knowledge to train the robot to master a new domain learning to avoid local minimum due to its access to only the partial information (Abbeel & Ng, 2004). Moreover, experts can also provide models for the RL robot to learn by itself (Hester & Stone, 2012).

This study addressed the current learning analytics models' challenges and proposed solutions for precision education. Thus, from a practical perspective, the theoretical advances made the proposed LA2.0 possible for precision education by allowing the brain component with the RL algorithm and statistics model to take in multimodal student data streams from its subsuming components for real-time policy establishment and gain the maximum reward based on a series of actions. Subsequently, automated feedback, or scaffolds can be supplemented to the students depending on their individual learning needs, through the RL's dynamic task creation, heterogeneous task deployment, and arbitrary data flow dependencies. After that, the lack of internal validity and immediacy issues can be resolved. The brain component collects learner characteristics data from multiple sources (e.g., psychological assessments, observations, and conversation log) across platforms and creates dynamic and heterogeneous tasks for the RL's self-learning. As a result, the learning analytics model can be generalizable to different samples using various criteria. Besides, the RL mechanism can fuse and respond to multiple sensory data from various input streams. Thus, learner data from one platform can be transferred to another to consider the personalized feedback generation and adaptive course recommendation, targeting each individual for goal-setting and self-regulation.

The learning analytics results are displayed via the dashboard component about learners' learning progress and that of the machine with sensible feedback and suggestions for achieving the learner and the machine's best learning performance. By recognizing and supporting the varied challenges of the current learning analytics model, the proposed integrative LA2.0 framework can fulfill LA's potential to provide personalized and just-in-time high-quality learning for precision education.

This study provided a promising integrative framework of human-and-machine symbiotic learning to inform the next-generation learning analytics models with theoretical and empirical supports. As a caveat, issues such as data format compatibility, data privacy, and information security may cause additional challenges for educational data mining due to significant differences in data privacy and information sharing mechanisms between data producers and data consumers (Wu et al., 2014). Aside from those, the proposed framework can address the challenges of the current learning analytics models to support precision education.

Acknowledgement

This research was partially supported by Grant 108-2511-H-009 -013 -MY3 from the Ministry of Science and Technology, Taiwan.

References

Abbeel, P., & Ng, A. Y. (2004). Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the Twenty-First International Conference on Machine Learning*, 1. doi:10.1145/1015330.1015430

Aljohani, N. R., Daud, A., Abbasi, R. A., Alowibdi, J. S., Basheri, M., & Aslam, M. A. (2019). An integrated framework for course adapted student learning analytics dashboard. *Computers in Human Behavior, 92*, 679–690. doi:10.1016/j.chb.2018.03.035

Baker, R. S. (2019, July). *Some challenges for the next 18 years of learning analytics*. Paper presented at the 9th International Learning Analytics & Knowledge Conference, Tempe, AZ.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic web. *Scientific American, 284*(5), 34–43. doi:10.1007/978-0-387-30440-3_478

Bettinger, E. P., & Long, B. T. (2018). Mass instruction or higher learning? The impact of college class size on student retention and graduation. *Education Finance and Policy, 13*(1), 97–118. doi:10.1162/edfp_a_00221

Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics, 3*(2), 220–238.

Cambria, E., Poria, S., Gelbukh, A., & Thelwall, M. (2017). Sentiment analysis is a big suitcase. *IEEE Intelligent Systems, 32*(6), 74–80. doi:10.1109/MIS.2017.4531228

Charleer, S., Moere, A. V., Klerkx, J., Verbert, K., & De Laet, T. (2018). Learning analytics dashboards to support adviser-student dialogue. *IEEE Transactions on Learning Technologies, 11*(3), 389–399. doi:10.1109/TLT.2017.2720670

Cook, C. R., Kilgus, S. P., & Burns, M. K. (2018). Advancing the science and practice of precision education to enhance student outcomes. *Journal of School Psychology, 66*, 4–10. doi:10.1016/j.jsp.2017.11.004

Dabbagh, N., & Kitsantas, A. (2012). Personal learning environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning. *The Internet and Higher Education, 15*(1), 3–8. doi:10.1016/j.iheduc.2011.06.002

Deshmukh, S. P., Patwardhan, M. S., & Mahajan, A. R. (2018). Feedback based real time facial and head gesture recognition for e-learning system. In *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data* (pp. 360–363). doi:10.1145/3152494.3167991

Ferguson, R., & Shum, S. B. (2012). Social learning analytics: Five approaches. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 23–33). doi:10.1145/2330601.2330616

Gamberini, L., & Spagnolli, A. (2016). Towards a definition of symbiotic relations between humans and machines. In *International Workshop on Symbiotic Interaction* (pp. 1–4). Cham, Switzerland: Springer. doi:10.1007/978-3-319-57753-1_1

Gerstein, J. (2014). Moving from education 1.0 through education 2.0 towards education 3.0. In *Experiences in Self-Determined Learning* (pp. 83-98). Scotts Valley, CA: CreateSpace Independent Publishing Platform.

Grafsgaard, J., Wiggins, J. B., Boyer, K. E., Wiebe, E. N., & Lester, J. (2013, July). *Automatically recognizing facial expression: Predicting engagement and frustration*. Paper presented at Educational Data Mining 2013, Memphis, TN.

Hester, T., & Stone, P. (2012). Learning and using models. In *Reinforcement learning* (pp. 111–141). Springer. doi:10.1007/978-3-642-27645-3_4

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255–260. doi:10.1126/science.aaa8415

Kar, N., Debbarma, M. K., Saha, A., & Pal, D. R. (2012). Study of implementing automated attendance system using face recognition technique. *International Journal of Computer and Communication Engineering, 1*(2), 100-103. doi:0.7763/IJCCE.2012.V1.28

Kitto, K., Lupton, M., Davis, K., & Waters, Z. (2017). Designing for student-facing learning analytics. *Australasian Journal of Educational Technology, 33*(5), 152–168. doi:10.14742/ajet.3607

Larmuseau, C., Vanneste, P., Desmet, P., & Depaepe, F. (2019). Multichannel data for understanding cognitive affordances during complex problem solving. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 61–70). doi:10.1145/3303772.3303778

Lee, Y.-H. (2018). Internet-based epistemic beliefs, engagement in online activities, and intention for constructivist ICT integration among pre-service teachers. *Australasian Journal of Educational Technology, 34*, 120–134. doi:10.14742/ajet.3747

Lee, Y.-H. (2021). Scaffolding university students' epistemic cognition during multimodal multiple-document reading: The effects of the epistemic prompting and the automated reflection report. *The Internet and Higher Education, 49,* 100777. doi:10.1016/j.iheduc.2020.100777

Lee, Y.-H., & Wu, J.-Y. (2013). The indirect effects of online social entertainment and information seeking activities on reading literacy. *Computers & Education, 67*, 168–177. doi:10.1016/j.compedu.2013.03.001

Lewandowska, M., Rumiński, J., Kocejko, T., & Nowak, J. (2011). Measuring pulse rate with a webcam – A Non-contact method for evaluating cardiac activity. In *2011 Federated Conference on Computer Science and Information systems (FedCSIS)* (pp. 405-410). Szczecin, Poland: IEEE.

Nishihara, R., Moritz, P., Wang, S., Tumanov, A., Paul, W., Schleier-Smith, J., Liaw, R., Niknami, M., Jordan, M. I., & Stoica, I. (2017). Real-time machine learning: The missing pieces. In *Proceedings of the 16th Workshop on Hot Topics in Operating Systems* (pp. 106–110). doi:10.1145/3102980.3102998

Pereira, J. (2016). Leveraging chatbots to improve self-guided learning through conversational quizzes. In *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 911–918). doi: 10.1145/3012430.3012625

Pinto, L., Davidson, J., Sukthankar, R., & Gupta, A. (2017). Robust adversarial reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning* (Vol. 70, pp. 2817–2826). Retrieved from https://arxiv.org/abs/1703.02702

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery, 10*(3), e1355. doi:10.1002/widm.1355

Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior, 78*, 397–407. doi:10.1016/j.chb.2017.06.030

Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S., & Kirschner, P. A. (2018). Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior, 107,* 105512. doi:10.1016/j.chb.2018.05.004

Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal, 3*(5), 637–646. doi:10.1109/JIOT.2016.2579198

Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning, 2*(1), 3–10.

Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 252–254). doi:10.1145/2330601.2330661

Spector, J. M. (2014). Conceptualizing the emerging field of smart learning environments. *Smart Learning Environments, 1*(1), 2. doi:0.1186/s40561-014-0002-7

Stowell, J. R., Oldham, T., & Bennett, D. (2010). Using student response systems (“clickers”) to combat conformity and shyness. *Teaching of Psychology, 37*(2), 135–140. doi:10.1080/00986281003626631

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). Cambridge, MA: MIT Press.

van Otterlo, M., & Wiering, M. (2012). Reinforcement learning and Markov decision processes. In *Reinforcement Learning Adaptation, Learning, and Optimization* (Vol. 12, pp. 3–42). Springer. doi:10.1007/978-3-642-27645-3_1

Vellido, A., Martín-Guerrero, J. D., & Lisboa, P. J. (2012). Making machine learning models interpretable. In *Proceedings of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)* (pp. 163–172). Retrieved from http://www.i6doc.com/en/livre/?GCOI=28001100967420

Vygotsky, L. S. (1987). *The collected works of LS Vygotsky: Volume 1: Problems of general psychology, including the volume Thinking and Speech* (Vol. 1). Springer. doi:10.1086/356870

Wu, J.-Y. (Online first). The predictive validities of individual working-memory capacity profiles and note-taking strategies on online search performance. *Journal of Computer Assisted Learning, 36*(6), 876-889. doi:10.1111/jcal.12441

Wu, J.-Y. (2014). Gender differences in online reading engagement, metacognitive strategies, navigation skills and reading literacy. *Journal of Computer Assisted Learning, 30*(3), 252–271. doi:10.1111/jcal.12054

Wu, J.-Y. (2015). University students’ motivated attention and use of regulation strategies on social media. *Computers & Education, 89*, 75–90. doi:10.1016/j.compedu.2015.08.016

Wu, J.-Y. (2017). The indirect relationship of media multitasking self-efficacy on learning performance within the personal learning environment: Implications from the mechanism of perceived attention problems and self-regulation strategies. *Computers & Education, 106*, 56–72. doi:10.1016/j.compedu.2016.10.010

Wu, J.-Y. (2020). Learning analytics on structured and unstructured heterogeneous data sources: Perspectives from procrastination, help-seeking, and Machine-Learning defined cognitive engagement. *Computers & Education*, *163*, 104066. doi:10.1016/j.compedu.2020.104066

Wu, J.-Y., & Cheng, T. (2019). Who is better adapted in learning online within the personal learning environment? Relating gender differences in cognitive attention networks to digital distraction. *Computers & Education, 128*, 312–329. doi:10.1016/j.compedu.2018.08.016

Wu, J.-Y., & Nian, M.-W. (2021). The dynamics of an online learning community in a hybrid statistics classroom over time: Implications for the question-oriented problem-solving course design with the social network analysis approach. *Computers & Education*, 104120. https://doi.org/10.1016/j.compedu.2020.104120

Wu, J.-Y., Hsiao, Y.-C., & Nian, M.-W. (2020). Using supervised machine learning on large-scale online forums to classify course-related Facebook messages in predicting learning achievement within the personal learning environment. *Interactive Learning Environments, 28*(1), 65–80. doi:10.1080/10494820.2018.1515085

Wu, J.-Y., & Peng, Y.-C. (2017). The modality effect on reading literacy: Perspectives from students’ online reading habits, cognitive and metacognitive strategies, and web navigation skills across regions. *Interactive Learning Environments, 25*(7), 859–876. doi:10.1080/10494820.2016.1224251

Wu, J.-Y., & Xie, C. (2018). Using time pressure and note-taking to prevent digital distraction behavior and enhance online search performance: Perspectives from the load theory of attention and cognitive control. *Computers in Human Behavior, 88*, 244–254. doi:10.1016/j.chb.2018.07.008

Wu, X., Zhu, X., Wu, G.-Q., & Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering, 26*(1), 97–107. doi:10.1109/TKDE.2013.109

Wu, Y.-C. J., Wu, T., & Li, Y. (2019). Impact of using classroom response systems on students’ entrepreneurship learning experience. *Computers in Human Behavior, 92*, 634–645. doi:10.1016/j.chb.2017.08.013

Yang, S. J. H. (2019). Precision education: New challenges for AI in education [conference keynote]. In *Proceedings of the 27th International Conference on Computers in Education (ICCE)* (pp. XXVII-XXVIII). Kenting, Taiwan: Asia-Pacific Society for Computers in Education (APSCE).

Yang, T.-C., Liu, Y.-L., & Wang, L.-C. (2021). Using an Institutional Research Perspective to Predict Undergraduate Students' Career Decisions in the Practice of Precision Education. *Educational Technology & Society*, *24* (1), 280-296.

Using an Institutional Research Perspective to Predict Undergraduate Students' Career Decisions in the Practice of Precision Education

Tzu-Chi Yang[1,2], Yih-Lan Liu[2] and Li-Chun Wang[3*]

[1]Department of Mathematics and Information Education, National Taipei University of Education, Taiwan (R.O.C.)// [2]Institute of Education/Center for Teacher Education, National Chiao Tung University, Taiwan (R.O.C.) // [3]Department of Electrical Engineering, National Chiao Tung University, Taiwan (R.O.C.) // tcyang.academic@gmail.com // ylliu@mail.nctu.edu.tw // lichun@g2.nctu.edu.tw

[*]Corresponding author

ABSTRACT: The recently increased importance of practicing precision education has attracted much attention. To better understand students' learning and the relationship between their individual differences and learning outcomes, the bird-eye view possible for educational policymakers and stakeholders from educational data mining and institutional research has gained importance and momentum. The deployment of specific predictive tasks based on institutional data is the most promising solution for dealing with a variety of issues on precision education. Most research in this field is focused on learning performance and related issues, such as at-risk students and drop-out tendencies. Seldom are the relationships between the learning performance and career decisions of students investigated. However, developing a deep understanding of students' career decisions plays an important role in the practice of precision education. In this vein, this paper provides a comprehensive analysis and comparison of the state of the art of predictive techniques for providing a prediction for students' career decisions. The results indicated that it is possible to perform early detection of students' career decisions. The contributions of this study are discussed in terms of their implications for theory, methodology, and application.

Keywords: Precision education, Institutional Research, Students' life planning, Machine learning, Educational data mining

1. Introduction

Due to the massification and diversification of higher education, ensuring broad student success has become a multifaceted challenge for higher education institutions (Arthars et al., 2019). Many researchers (Chen and Wang, 2020) have noted the importance of implementing personalized learning to accommodate students' individual differences, as such differences are closely related to learning performance, attitudes, career decisions, self-efficacy, and success in learning. The provision of learning support without loss of timeliness or personalization is promising (Kokkinos & Saripanidis, 2017; Pardo, 2018; Valladares et al., 2018). The implementation of personalized learning relies on the collection and analysis of learning data drawn from a range of students (Beemer et al., 2018). These records, however, are commonly available in institutional file systems/databases and independently stored in different departments. An institutional database that can store a range of data related to students' individual characteristics and learning is valuable for conducting institution-specific research under conditions of resource constraint (Caison, 2007).

Likewise, better or more comprehensive learning data collection and exploration in relation to learning, from an institutional perspective, has been proposed to support personalized learning (Arthars et al., 2019; Flaig, Simonsmeier et al., 2018). This is because data are easy to find in institutions and can be incorporated to support decision-making (Swing & Ross, 2016). Understanding the rationales and applications of this approach can also be furthered through an objective and reflective examination of historical trends, current practices, and the strengths and weaknesses of teaching and training strategies (Ross & Morrison, 2012). In this vein, the use of an institutional database to deal with personalized learning issues has attracted much attention (Camacho & Legare, 2016; Laguilles, 2016). Education as a whole is also showing progressive movement from a one-size-fits-all approach to a type of precision or personalized education (Lu et al., 2018; Tsai et al., 2020).

Researchers have argued that most personalized learning solutions, in the form of educational tools and flexible learning systems, accommodate individual learners' interactions and learning progress, and they fit the specific needs of the individual learner (Xie et al., 2019; Zawacki-Richter et al., 2019). Nevertheless, data analytics have become inseparable from policy development and governance modes. There is a need to take up new knowledge and provide advanced resources in education that are based on data-led policymaking, such as data-driven precision education (Gulson & Webb, 2017; Williamson, 2019).

Precision education is used to identify students' personal characteristics and individual needs to provide educational researchers and practitioners with the tools to understand students better and to allow for a more effective approach to education (Hart, 2016; Makhluf, 2020). For example, Rump et al. (2017) predicted students' intention to drop out based on learning motivation. Their results indicated that higher education institutions can adopt prophylactic measures that could prevent college students from dropout. Hoffait and Schyns (2017) developed institutional data mining to early detect freshmen likely to face difficulties to allow universities to timely and efficiently provide remediation or reorientation. Cheng et al. (2018) predicted potential high-risk freshmen in three core university courses using institutional databases and demonstrated that this predictive task can ameliorate the problems of freshman unpreparedness by providing intervention with the support of the Office of Student Affairs.

In this light, researchers (Thompson et al., 2019) have investigated the myriad of stressors that students face as they transit into university study, during which they continue to balance competing demands of academic study, social relationships, and personal needs. Researchers have also found that many students are concerned about their educational and career decisions during their college experience (Miller et al., 2018). Hence, greater attention than ever is being placed on the ways in which universities enable their graduates to meet their career goals (Healy et al., 2020). Thus, students' career decisions should be taken into account as they develop strategies for precision education.

On the one hand, students rely on the learning support and guidance provided by universities, especially as relates to career path development. On the other hand, allocating these learning resources to students relies on supportive educational policy making and practice that is associated with institutional research. Such institutional methods must be examined through a bird's eye view. To this end, this study proposes a framework that uses historical trends to predict undergraduate students' career decisions. More specifically, it explores the relationships among individual students' characteristics, learning performance, and career decisions in relation to institutional data in pursuit of this research question: "Is it possible to predict students' career decisions based on individual differences and learning data?" Then, the second research question, namely, "How can the career decisions predictive model can enhance the practice of precision education?" is addressed within the context of precision education for institutional research.

2. Literature review

2.1. Precision education

Precision education can be defined as an approach to research and practice that allow preventive and intervention practices to be tailored to individuals based on the best available evidence. Put another way, it is an approach to creating a learning environment that provides learners with precise instruction, assistance, and resources according to their individual needs, based on the best available evidence (Cook et al., 2018). This approach is inspired by precision medicine, as mentioned by President Barack Obama (Collins & Varmus, 2015; Hart, 2016). Precision medicine focuses on understanding individual variability in disease prevention, care, and treatment in a way that takes into account individual variability in genes, environment, and lifestyle (Collins et al., 2016; Reed & Gates, 2020). In relation to the educational context, the interventions that are most aligned with learners' needs can be precisely evaluated and delivered by linking particular issues (e.g., social, academic, and physical health) to create personalized educational experiences for learners (Burns et al., 2010; Palanica et al., 2019). This approach has attracted much attention and is expected to produce superior outcomes to standardized one-size-fits-all or trial-and-error approaches that are often seen in education (Cook et al., 2018; Gulson & Webb, 2018).

For examples, Bahr (2009) explored the relationship between occurrence and frequency of students' lateral transfer in relation to their completion of a credential and he classified students' behavior pattern into six groups (e.g., vocational, drop-in, noncredit, and exploratory). Prediction of student behavior pattern allows learning resource arrangements to be made to improve long-term outcomes (Bahr, 2010). Laanan and Jain (2016) indicated that the transfer process can be a daunting experience for students who aspire to complete a bachelor's degree. Institution research plays an important role in helping educational institutions evaluate complex processes so that relevant strategies can be developed to deal with issues that arise. Lu et al. (2018) considered precision education as a framework for use in improving learning performance. They collected and analyzed students' learning behaviors and to predict their learning performance from them. From this, certain critical features, such as online learning (e.g., video-viewing and out-of-class practice), homework and quiz scores, and after-school tutoring were found to be closely related to learning performance. Accordingly, these researchers suggested that monitoring and intervening in students' online and off-line learning behaviors in response to their observed learning behavior patterns could be a promising approach to improving learning performance and practicing precision education. Likewise, O'Connor and Daly (2018) examined different combinations of antecedent- and consequent-based strategies for students with escape-motivated problems. They found that students with escape-motivated behavior are a heterogeneous group. This implies that it is possible to develop individualized solutions (e.g., teaching, consequent strategies) to a target student.

Following the rationale of precision education, researchers (Cook et al., 2018) developed the idea that the practice of precision education, including (a) flexible and adaptive interventions that allow for adaptation and matching to student needs, (b) formatively monitoring and evaluation of data over time that allows stakeholders to further modify interventions given sufficient response or to select an alternative strategy. Then, (c) this amounts to collaboration, allowing the making of timely and appropriate decisions regarding intervention adaptation and matching. This might modify the existing intervention, and implement an alternative one, or simply terminate the intervention outright.

In a nutshell, the practice of precision education involves tailoring mechanisms of prevention and intervention to certain individuals, following the best evidence regarding their needs and potential, established with evidence-based approaches. This allows the selection and provision of more precise treatments for individual learners. This not only enhances learning outcomes but also avoids educational waste, such as the costs in time, finances, and resources (Kazdin & Blasé, 2011).

2.2. Institutional research perspectives on precision education

As the aforementioned, precision education refers to evidence-based innovation in education that takes advantage of big data and data science to enhance traditional educational practices (Christensen & Eyring 2011; Tsai et al., 2020). The critical process here is to analyze big data and develop models to predict student success, meet individual needs, and fulfill educational purposes, such as academic performance, graduation rates, resource consumption, counseling, and many others (Warren et al., 2017). Comprehensive data sources of individuals' characteristics, learning contexts, and portfolios are typically organized and conducted in data/information centers with institutional research professionals (Lonn & Koester, 2019; Rienties & Toetenel, 2016). Further, precision education is a part of a rising uptake of data science and domain knowledge in educational policy, institutional research, and advocacy for data-led policymaking (Gulson & Webb, 2017; Williamson, 2019). Many researchers have argued that it is particularly important to collect, organize, analyze, and utilize comprehensive learning data to support making decisions from a bird's-eye view (e.g., institutional research) to achieve precision education (Bellgard, 2020; Kuch et al., 2020).

Over the past decade, various research approaches and frameworks have been proposed for the practice of precision education in an institutional research perspective (Xing et al., 2015). For example, Márquez-Vera et. al. (2016) proposed a methodology for creating early prediction models for student dropout as soon as possible. They found that this method was sufficiently trustworthy to be used in an early warning system to improve student retention. Du et al. (2019) developed a framework for predicting the learning performance and creating an early warning system for at-risk students to improve learning outcomes. Adekitan and Salau (2019) explored the impact of engineering students' performance in the first three years on their graduation results. If students' learning performance can be predicted, early intervention can be deployed to prevent them from graduating with poor results or from not graduating at all.

Early prediction provides various opportunities to assess information on student learning and to facilitate decision making to serve improvements in education (Daghestani et al., 2020). Most research into this field focuses on learning performance, although student learning could be influenced by many factors, such as career factors (Freeman et al., 2017; Kim et al., 2018). Researchers have found that people seek to become employable by pursuing individually tailored precision education and learning based on their individual needs and what they hope to achieve in life (Santos et al., 2018; Brunila, 2020).

Recently, more attention has been given to producing appropriate methods and contents for career counseling in university students (Peng et al., 2020). This is because individual learning experiences and background factors affect self-efficacy, learning experience, and career path development in students. Students seek to rely on the resources provided by the university, such as learning facilities and career guidance, when they consider the academic or career path they are hoping to follow (Berger et al., 2019; Rivera & Li, 2020). Here, universities are increasingly aware that they have a major responsibility to assist students, not only in academic development but also in career development (Meijers & Kuijpers, 2014). Accommodating the individual need to make career decisions and providing sufficient support to students are an important goal in the practice of precision education (Abdullah et al., 2018; Ulas & Yildirim, 2019). Because precision education relies on educational policy making and practices that are associated with institutional research, it would be promising to explore the methods of practicing precision education in the university with a large-scale view of the university. Thus, we use an institutional research perspective to propose a model for predicting student career decisions and to elaborate how such a model can be used in the practice of precision education.

3. Method

3.1. Data source and data selection

The data used in this study were collected from a de-identified institutional data warehouse supervised by the Big data research center (BDRC) of National Chiao-Tung University in Taiwan. A data warehouse is the repository of the stored data of an institution and is designed to facilitate analysis, and reporting solutions that deal with institutional research issues. In general, it is a duplicate database system that dumps raw data from more than 80% of departments/centers in the university. Generally, the data warehouse supervised by BDRC of the aforementioned university are built with an Extract-Transform-Load (ETL). All data were pre-processed to remove all personally identifying information (e.g., name, student ID number, and contact information), following the privacy policy. A comma-separated-values (csv) file was generated and delivered to the researcher after the dataset application was complete. This study used a sample of 7003 undergraduate students who had enrolled in the university from August 2010 to September 2015, and who had received a bachelor's degree in no more than five years, and were not suspended or did not drop out.

3.2. Conceptual model of data pre-processing

The target university is a research university, known for its science and engineering achievements, especially in the fields of electronics, information communications, and optoelectronics. The university is located near to a Science Industrial Park. After graduation, most university students either seek an academic career or a position in industry. Therefore, in this study, career decisions were classified into four categories, i.e., academic career, engagement in advanced studies in current university, industry employment, and engagement in advanced studies at other universities.

To predict career decisions, various features beyond demographic data were considered, following previous research. For example, many researchers have shown that learning performance may affect career development in college students. Students with lower achievement might seek a different career from those with higher achievement. Thus, learning performance is worth considering as a factor in career decision (Chen, Chen, Hu, & Wang, 2015; Peng et al., 2020; Park & Lee, 2020). In addition to this, previous studies have found that mathematics knowledge and learning achievement in mathematics-related courses might be particularly closely associated with career planning in a college/university setting (Luzzo et al., 1999). This is because mathematics is a coherent discipline with concepts and topics that have important logical and conceptual connections. It is essential for students as they face challenges (problem) in life, reflecting the coherence of discipline. Students usually need to complete formal or informal mathematical education before engaging in any career path they choose (Mielicki et al., 2019). Mathematics knowledge is required for college and is associated with career readiness (Cogan et al., 2019). Accordingly, features that reflect mathematical ability (e.g., level of coursework, score on general scholastic ability test for mathematics, average scores for calculus-related courses) were taken into account in this study.

In brief, demographic data, learning performance, and mathematical knowledge could predict career decisions in the aggregate (Ferrare, & Miller, 2020; Ketenci et al., 2020; Surrette, 2020). The data used included demographic information, learning performance in university, and career decision, as shown in Figure 1. The demographic information was taken into account as a set of initial features prior to enrolling at the university. The career data included career information for each student after receiving their bachelor's degree for one year. Learning performance included the number of credits selected, the number of credits earned, and the average grade received each year. In addition, learning performance was divided into performance on calculus-related courses and on non-calculus-related courses.

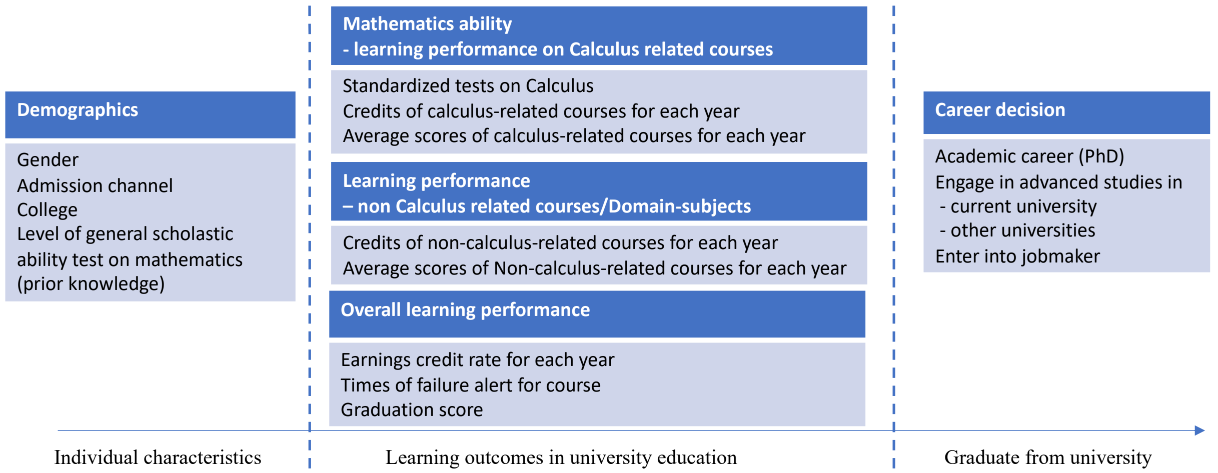


Figure 1. The conceptual model

Calculus-related courses (CRCs) were assessed by two assistants who have both BA and MA degrees in mathematics, and the assessment was validated by a professor who has one-year teaching experience in calculus. A course would be identified as a CRC if (a) calculus was considered basic knowledge in the course (e.g., differential or integral) or (b) if calculus was frequently used for solving problems in the course (e.g., Fourier transformation). Typical calculus-related courses included engineering economics, cryptography, and engineering mathematics. A course not identified as CRC would be classified as an NCRC. All newly recruited students were required to take a standardized test in calculus during the first two semesters, and so the CRC and NCRC values for the first year were replaced by the standardized tests scores for calculus. The overall learning outcomes, including the earnings credit rates for each year, times of failure alert for the course, and graduation score, were incorporated into analysis. All features were processed based on the conceptual model of the features, including calculating values and labelling for specific features. The criteria for labelling features and the schema for the analysed features are given in Tables 1 and 2, respectively. The criteria for labelling features and the schema for the analysed features were given in Tables 1 and 2, respectively.

Table 1. Descriptions of criteria of labelling feature

Feature	Description	Criteria of label	Replaced
Career	Academic career	• Enrolled in the MA program after receiving a BA degree and enrolled in the Ph.D. program after receiving an MA degree • Enrolled in the Five-Year BA-MA Program and Ph.D. program • Enrolled in the Five-Year BA-Ph.D. program • Received a BA degree for one or years and enrolled in the Ph.D. program after receiving an MA degree	AC
	Engage in advanced studies in current university	• Enrolled in the MA program after receiving a BA degree • Enrolled in the MA program after receiving the BA degree for more than one year • Enrolled in the Five-Year BA-MA Program	AS
	Employment	Do not enroll in any further study program and work in a full-time position (according to instance response to Graduation Questionnaire)	EM
	Engage in advanced studies in other universities	Responded that he/she had enrolled a further study program (e.g., MA, PhD.) in other universities	OT
GSATm	Highest level	The original value range of GSAT was a 0-15 ordinal scale, which 15 is the highest score. GSATm was labeled based on the Quartile of the value range obtained from the dataset in this study, i.e., A: GSATm > Q1 B: Q2 ≤ GSATm < Q1 C: Q 3 ≤ GSATm < Q4 D: GSATm ≤ Q4	A
	High level		B
	Middle level		C
	Lowest level		D

Table 2. Structure of the analysed data

Feature	Description	Type of data	Value range of feature
Demographic			
GSATm	Level of general scholastic ability test on mathematics	Ordinal	A: highest, B: high, C: medium, D: low
Gender	Gender of instance	Nominal	0: Female, 1: Male
AC	Admission channel	Nominal	AE: admission by examination and placement, PA: personal application, SP: star Program
College	College of instance	Nominal	ES: College of Science, EE: College of Electrical and Computer Engineering, CS: College of Computer Science, E: College of Engineering, M: College of Management, SS: College of Humanity and Social Science, BT: College of Biological Science and Technology, HS: College of Hakka Studies
Learning performance on Calculus			
STC1	1st Standardized test on Calculus	Numeric	0-100
STC2	2nd Standardized test on Calculus	Numeric	0-100
CRC2 CRC3 CRC4	Credits of calculus-related courses for 2nd/3rd/4th year	Numeric	0-N
CRCs2 CRCs3 CRCs4	Average scores of calculus-related courses for 2nd/3rd/4th year	Numeric	0-100
Learning performance on non-calculus subjects			
NCRC1 NCRC2 NCRC3 NCRC4	Credits of non-calculus-related courses for 1st/2nd/3rd/4th year	Numeric	0-N
NCRCs1 NCRCs2 NCRCs3 NCRCs4	Average scores of Non-calculus-related courses for 1st/2nd/3rd/4th year	Numeric	0-100
Overall learning performance			
ECR1 ECR2 ECR3 ECR4	The earnings credit rate for 1st/2nd/3rd/4th Year	Numeric	0-1
TFA	Times of failure alert for course	Numeric	0-N
GS	Graduation score	Numeric	0-100
Career decision			
CD	Individual career decision	Nominal	AC: Academic, AS: engage in advanced studies in current university, EM: Employment, OT: engage in advanced studies in another university

3.3. Dataset

As seen in the schema, 27 features were selected in this study, and each data row provides these features for a given student. However, no data rows that contained any missing values were taken into consideration and were excluded from the dataset. After this operation, 6255 data rows remained in the dataset.

Admission by examination and placement (AE), personal application (PA), and star program (SP) were three major admission channels that students could use to apply to university. The records we collected in this study only covered students recruited through these admission channels. SP was the latest admission channel launched in Taiwan, and it was used to recruit students from underrepresented high schools. SP provides opportunities for students with good rankings within their senior high schools but who had average uniform high school examination scores. SP is also developed as a means of promoting educational equality for all. The PA channel focuses on scores of nationwide General Scholastic Ability Test (GSAT) and on personal multiple performance at higher schools. That is, students recruited through this channel may have an excellent test scores and have a potential match to one or some of the foci of the universities. On the other hand, students who are not admitted to college via the SP and PA take the Advanced Subjects Test and are then allocated academic majors and universities according to their scores via the AE channel.

Table 3. Career decision of students from different admission channels

Admission channel	Career decision				
	AC	AS	EM	OT	Total
AE	71	805	1657	193	2730
PA	71	1039	1334	203	2649
SP	29	388	390	69	876
Total	171	2232	3381	465	6255

With this, it should be noted that students recruited through these admission channels may have different personal characteristics and backgrounds due to different criteria. Considering the possible heterogeneity of the students from these channels (Lin & Liou, 2019), we first evaluated the distribution (Table 3) by applying the chi-square test ($\chi^2(8) = 104.28$; $p < .001$), and significant differences were found between the career decisions of students from three admission channels. That is to say, students in the three groups exhibited some heterogeneity. For this reason, we generated sub-datasets according to admissions channel, i.e., dataset A: AE, dataset B: SP, and dataset C: SP, to further evaluate the performance of predictive modeling by comparing datasets that split by admission channels and those that did not.

3.4. Implementation of university students' career planning prediction

Many similar precedent studies have examined the relationships between career decisions and individual characteristics of students using conventional statistical models (Chuang & Dellmann-Jenkins, 2010). However, statistical models are usually designed to infer relationships between variables, and machine learning models could be used to maximize predictive accuracy (Tsai et al., 2020). Regarding decisions for practical precision education, recent studies have suggested that supervised machine learning approaches are the most often used to create predictive models for classification and prediction tasks, and these approaches were adopted here as well (Dutt et al., 2017; Rodrigues et al., 2018). Most approaches were model based, driven by an estimation of implicit correlation among analyzed samples and input features and comprising an underlying model that supports the performance of the predictions task. Typical approaches in this area include the artificial neural network (ANN), the support vector machine (SVM), multinomial logistic regression (MLR), decision tree (DT), and rule induction (RI).

In addition to a model-based approach, similarity-based approaches (e.g., k-nearest neighbor, KNN) and probabilistic approaches (e.g., naive Bayes, NB) could be adopted to obtain a comparison of models in a predictive task. The former constitutes the discovery of similar features in provided training data and evaluates the category of a test instance based on the similarity (e.g., the discovery of students which have similar GPA), while the latter exploits probability distribution characteristics that can be observed in training data. To obtain a comprehensive comparison of performance in different predictive models, the seven models described above were also taken into account.

3.5. Evaluation criteria

The confusion matrix was used to describe the performance of a classifier model. The table of a confusion matrix incorporates true predictions and prediction errors. Figure 2 presents a four-class prediction task {A, B, C, D}, where TPA can be defined as the number of observations classified as A where A was predicted, while EAB refers to the number of observations classified as B where A was predicted. The overall accuracy, or kappa (Cohen, 1960), precision, and sensitivity (also called recall) can be calculated to present a measurement for the evaluation of predictive models. The calculation criteria and details of the above measurements are described as follows:

	A	B	C	D
A	TP_A	E_{AB}	E_{AC}	E_{AD}
B	E_{BA}	TP_B	E_{BC}	E_{BD}
C	E_{CA}	E_{CB}	TP_C	E_{CD}
D	E_{DA}	E_{DB}	E_{DC}	TP_D

Figure 2. Illustration of the confusion matrix

- Overall Accuracy = $(TP_A + TP_B + TP_C + TP_D) / (TP_A + TP_B + TP_C + TP_D + E_{AB} + E_{AC} + E_{AC} + E_{BA} + E_{BC} + E_{BD} + E_{CA} + E_{CB} + E_{CD} + E_{DA} + E_{DB} + E_{DC})$
- Kappa = $(P_0 - P_e) / (1 - P_e)$
  where P_0 is the proportion of the observed agreements, P_e is the proportion of agreements expected by chance
- Precision $_{\text{Class A}}$ = $TP_A / (TP_A + E_{AB} + E_{AC} + E_{AD})$
- Sensitivity $_{\text{Class A}}$ = $TP_A / (TP_A + E_{BA} + E_{CA} + E_{DA})$

Here, overall accuracy describes how well a model predicts the career planning of all students investigated. This measure evaluates how often the predictive model is correct. Kappa compares the accuracy of the predictive model with that which a random classification is expected to achieve. The value range is less than or equal to 1, where 1 indicates a perfect prediction by the predictive model, and 0 indicates no better than a random guess. The kappa value is usually considered to measure the success of a predictive model in an imbalanced dataset and thus was taken into account in this study. Precision was defined as the fraction of correct predictions for a certain class. Sensitivity was defined to measure, where a certain class should have been predicted, how many times it was correctly predicted.

To identify the performance of each model for predicting university students' careers, accuracy and kappa were used and precision and sensitivity were used to further evaluate the performance of predictive models for different institutional research purposes.

4. Results and discussions

4.1. Overall performance of the predictive model

The results (Table 4) showed that for high consistency (kappa > 0.7), most models can achieve an accuracy of over 86.76%. That is to say, it is reasonable to expect accurate prediction of university students' career decisions from predictive models having the features proposed in this study. In institutional research, such technologies can be used to provide additional information for the competent authority and stakeholders to address issues related to precision education, such as the arrangement of educational resources and the development of specific strategies. For example, it can predict whether the number of students who choose to enter the job market directly after graduation from university. Educational resources, such as domain-subject curricula, career development instructors, and related programs can be evaluated to determine whether they are sufficient.

The predictive models used in the distinguished dataset demonstrated higher accuracy and consistency than those in the non-distinguished dataset. In other words, the distinguished dataset may produce better performance in predicting career decisions. This result echoed the results of a chi-square test that revealed heterogeneity among the three admissions channel groups. Hence, removing features of the admissions channel improved the accuracy and consistency of the predictive models. These results suggested a need to distinguish different admissions channels for the dataset when performing prediction tasks.

Based on the presented results, it was noticed that by feeding data dataset A (channel: AE) and dataset B (channel: PA), the best performance was obtained by using DT followed by ANN. For feeding dataset C (channel: SP), the best performance was obtained by using ANN followed by DT. That is to say the use of DT and ANN predictive models achieved the highest performance in this prediction task. Most university students in Taiwan were recruited from these three admission channels. This result also suggested that DT and ANN were worth being taken to deal with the career decision prediction tasks.

Table 4. Classifier prediction accuracy and kappa (sorted by accuracy)

Dataset split by admission channel						Original dataset	
Dataset 1 (AE)		Dataset 2 (PA)		Dataset 3 (SP)		Dataset 4 (all channel)	
Classifier	Accuracy/ Kappa	Classifier	Accuracy	Classifier	Accuracy	Classifier	Accuracy
Decision Tree	86.76% / 0.737	**Decision Tree**	86.36% / 0.76	**ANN**	94.75% / 0.911	**Decision Tree**	83.17% / 0.6901
ANN	83.57% / 0.677	**ANN**	78.69% / 0.608	**Decision Tree**	82.99% / 0.703	**ANN**	75.93% / 0.5497
KNN	65.92% / 0.388	KNN	64.34% / 0.373	KNN	72.49% / 0.528	KNN	67.83% / 0.431
Rule Induction	70.10% / 0.355	Rule Induction	66.23% / 0.363	LogisticReg.	68.26% / 0.4385	LogisticReg.	67.05% / 0.3793
LogisticReg.	68.42% / 0.337	LogisticReg.	64.75% / 0.357	Rule Induction	68.26% / 0.4292	Rule Induction	67.63% / 0.3715
SVM	68.12% / 0.309	SVM	64.53% / 0.355	SVM	65.41% / 0.3781	SVM	66.73% / 0.3711
NaiveBayes	49.67% / 0.242	NaiveBayes	43.07% / 0.196	NaiveBayes	42.69% / 0.2309	NaiveBayes	48.46% / 0.2338

Classifier prediction accuracy for DT and ANN over a four-year period was also examined. As shown in Figure 3, increased accuracy was obtained in the third year, and the highest accuracy was found in the fourth year. The accuracy of predictions was evaluated using 10-fold cross-validation, as shown in Table 5. The prediction sought to provide precise learning resources and career guidance to students as earlier as possible, as the earlier the career decision of a student can be identified, the better the support that can be provided. These results indicated that performing such prediction tasks in the third year might be an appropriate time.

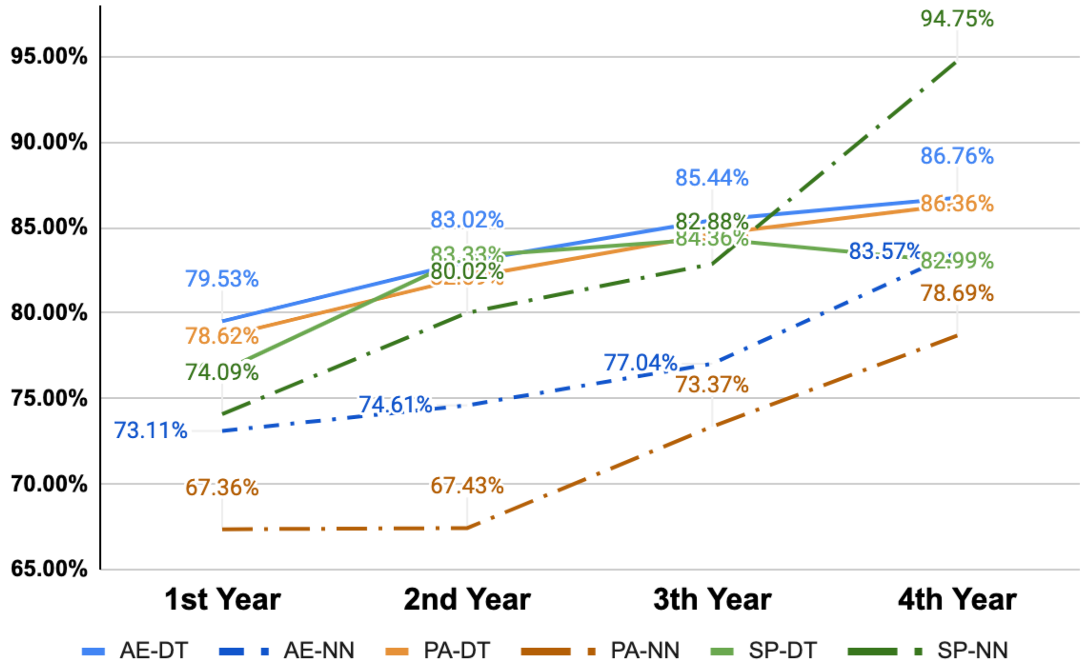


Figure 3. Classifier prediction accuracy of DT and ANN in four years

Table 5. Cross-validation results for DT and NN in each year

		1st year			2nd year			3rd year			4th year		
		AE	PA	SP	AE	PA	SP	AE	PA	SP	AE	PA	SP
DT	Accuracy	62.88%	56.93%	58.33%	62.62%	58.52%	59.36%	64.16%	57.91%	60.27%	64.09%	56.97%	59.70%
	Kappa	0.2451	0.2297	0.2806	0.2526	0.2584	0.2955	0.2703	0.256	0.3122	0.2763	0.2447	0.3026
NN	Accuracy	64.45%	60.94%	58.45%	64.23%	60.48%	55.37%	63.90%	61.69%	56.39%	64.75%	61.16%	59.13%
	Kappa	0.2736	0.2813	0.269	0.2843	0.2776	0.2142	0.2674	0.304	0.2423	0.2961	0.2922	0.2966

4.2. Precision and Sensitivity in university students' career decisions

Moreover, we further explored the precision and sensitivity of DT and ANN. Table 6 represents the confusion matrix for DT and ANN in predicting students' career decisions. Because this is a four-class classification problem, precision and sensitivity were calculated separately for each of the classes.

The precision of the prediction model indicates the rate at which a class was correctly predicted, and the sensitivity indicates that given a certain class should have been predicted, how many times it actually was correctly predicted. The overall precision for each class ranged from 0.5 to as high as 0.96, while sensitivity ranged from 0.4 to 0.97; however, in some classes, the models appeared to be inadequate or unable to provide predictions.

With the use of ANN, the precision and sensitivity of class AC was 0 for AE and PA students. This suggests that ANN could fail to identify students with a tendency toward enrolment in the PhD program if they were recruited from the PA or AE channel. In other words, educators and stakeholders might need to select a predictive model that is more appropriate for the goal of the prediction task.

Table 6. Confusion matrices for each entrance channel

Decision tree							ANN						
Channel = AE		Actual				Class precision	Channel = AE		Actual				Class precision
		AC	AS	EM	OT				AC	AS	EM	OT	
Predicted	AC	37	12	16	6	**0.889**	Predicted	AC	0	35	31	5	0
	AS	1	654	145	5	0.86		AS	0	693	108	4	0.775
	EM	4	53	1596	4	0.839		EM	0	105	1537	15	0.874
	OT	0	17	98	78	0.881		OT	1	61	83	48	0.667
Class sensitivity		**0.521**	0.812	0.963	0.404		Class sensitivity		0	**0.861**	0.928	0.249	
Channel = PA		Actual				Class precision	Channel = PA		Actual				Class precision
		AC	AS	EM	OT				AC	AS	EM	OT	
Predicted	AC	37	22	12	0	**0.902**	Predicted	AC	0	39	32	1	0
	AS	1	923	107	8	0.841		AS	0	873	165	1	0.778
	EM	2	106	1223	3	0.877		EM	0	125	1209	0	0.794
	OT	1	46	53	103	0.904		OT	117	85	117	1	0.5
Class sensitivity		0.521	0.888	**0.917**	0.507		Class sensitivity		0	0.84	**0.906**	0.005	
Channel = SP		Actual				Class precision	Channel = SP		Actual				Class precision
		AC	AS	EM	OT				AC	AS	EM	OT	
Predicted	AC	8	14	5	2	**0.889**	Predicted	AC	15	6	4	4	0.833
	AS	1	341	45	1	0.859		AS	2	383	3	0	0.946
	EM	0	25	361	4	0.84		EM	1	9	378	54	**0.964**
	OT	0	17	19	33	0.825		OT	1	7	7	54	0.885
Class sensitivity		0.276	0.879	**0.926**	0.478		Class sensitivity		**0.517**	0.986	**0.969**	0.783	

The measures of precision and sensitivity in particular provide valuable information for stakeholders as they select predictive models to deal with questions regarding precision education in relation to institutional research. Many researchers have noted that as students consider which academic or career path they would like to take, they rely on the resources and guidance provided by schools/colleges (Schwartz et al., 2016; Xie & Reider, 2014). More and more, universities are acknowledging that their strong responsibility to guide and support students as they begin their career development (Meijers & Kuijpers, 2014). Thus, correctly predicting career decision making could help educational institutions provide relevant support and resources to help students develop their career plans.

For example, suppose that a university is seeking to provide adequate employee training resources for students (e.g., practical courses, career consulting). A high accurate prediction for students who tend to enter the job market can ensure appropriate delivery of relevant resources and promotions. If this is ensured, career development resources can enjoy maximum utilization.

For a relevant example, supposed a university is seeking to enhance its PhD programs and provide adequate early training resources for students who look for an academic career by enrolling in a PhD program. The more students with such tendencies that can be identified, the better and more effective the early training will become.

In this vein, choosing a model with high sensitivity is particularly appropriate. In a nutshell, high accuracy (greater than 80%) was obtained for most predictive models. DT and ANN demonstrated the highest performance in predicting the career decisions of university students. In relation to precision education, both precision and sensitivity should be considered to help stakeholders choose appropriate predictive models for strategy development.

Conduct-appropriate predictive models might not have the highest contribution. More information related to students' career decisions is required. ANN simulates human nerves in a black-box manner, and so its predictive process is difficult to explain. Thus, little information was provided from this model. However, a decision tree was constructed by calculating estimated measurements (e.g., information gain or the Gini index) among analyzed features. Accordingly, the importance of each feature, classification rule, and predictive process in the DT model was explored. Thanks to this, using the DT model can easily provide a visualized result, namely, a tree-structured graph. This result can not only help educational decision-makers and stakeholders understand the predictive process but also help them extract information from its results (Ellis, 2019). The following section explains how to read and draw out information from DTs.

4.3. Classification rules and information extraction from decision trees

A decision tree is a series of nodes, starting at the base with a single node (root), passing through decision nodes and extending to terminal nodes (leaf nodes) that represent the categories that the tree can classify. It works like a flow chart, choosing a path at each split until a decision is ultimately made at the leaf nodes. These rules provide a detailed explanation for classification queries and extract useful information from DTs. For example, Figure 4 represents the predictive process for dataset A, and we can derive the following rules, given a student recruited by admission examination and placement.

R01: IF [ECRate 2 <= 79.54%], THEN CSL = EM

R02: IF [ECRate 2 <= 79.54%] AND [College = SS, HS], THEN CSL = EM

R03: IF [ECRate 2 <= 79.54%] AND [College = CS] AND [GS<=79.05], THEN CSL = EM

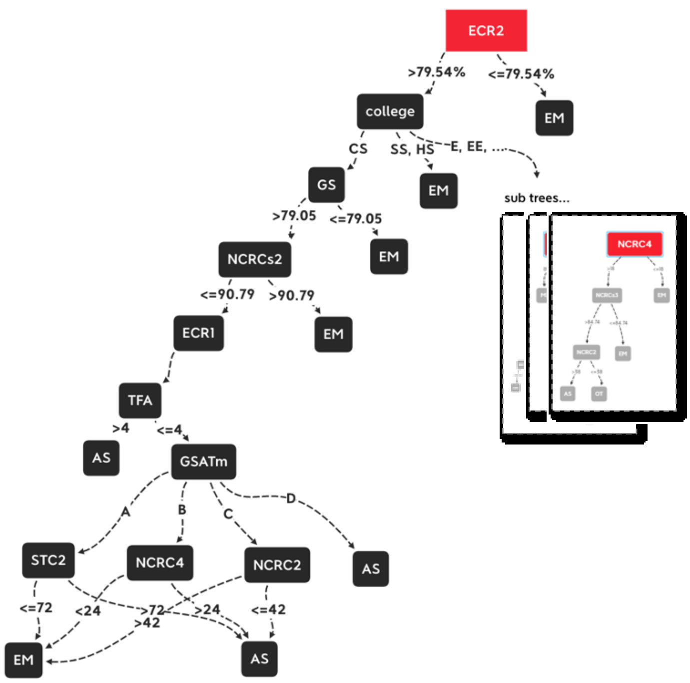


Figure 4. An example of decision tree in predicting career decisions of AE students

Traveling from the root down single or multiple paths, we can break down the tree into smaller and smaller subsets, incremental travel subsets with decision nodes and leaf nodes, to develop a final result. Each decision node has two or more branches and represents a test. The leaf nodes represent predictions (or classifications) for the paths. In this study, the predictions were academic (AC), engaging in advanced study at the current university (AS), employment (EM), and engaging in advanced studies at another university (OT).

Valuable information can be explored in this process. For instance, as shown in Figure 5, the root decision node for ECR2 has two branches, with one greater than 79.54% and ECR2 equal to or less than 79.54%, the student will tend to enter the job market after receiving the BA degree. Traveling down the decision tree for the other two channels (PA and SP), it appears that a student recruited from channel PA or SP tended to enter the job market if their the ECR3 was lower than 91.18% or ECR2 was lower than 72.73%, respectively. These results suggest that the rate at which university students earned credits plays an important role in their career decisions, in particular their rates during their second and third years. Additionally, most students in the College of Social Science (SS) and Hakka Studies (HS) were observed to enter the job market regardless of demographic and learning performance features (e.g., gender, level of general scholastic ability test on mathematics, rate of earning credits, or graduation score).

With regards to institutional research, these results indicated that students who tended to enter the job market after receiving a BA probably demonstrated lower rates of earning credits during the second and third years, except for those in SS and HS. In this vein, an early investigation can be performed to develop a deeper understanding of students' career planning. In this way, institutions can arrange and provide adequate resources for students to help them develop their career plans and achieve their goals.

Likewise, the time of failure alert (TFA), level of general scholastic ability test on mathematics (GSATm), and non-calculus-related credit in the fourth year (NCRC4) for students are useful statistics to identify students who would will seek to apply to a master's program, regardless of admission channels. Accordingly, more academic training resources (e.g., advanced courses for domain subjects) can be arranged and presented to students.

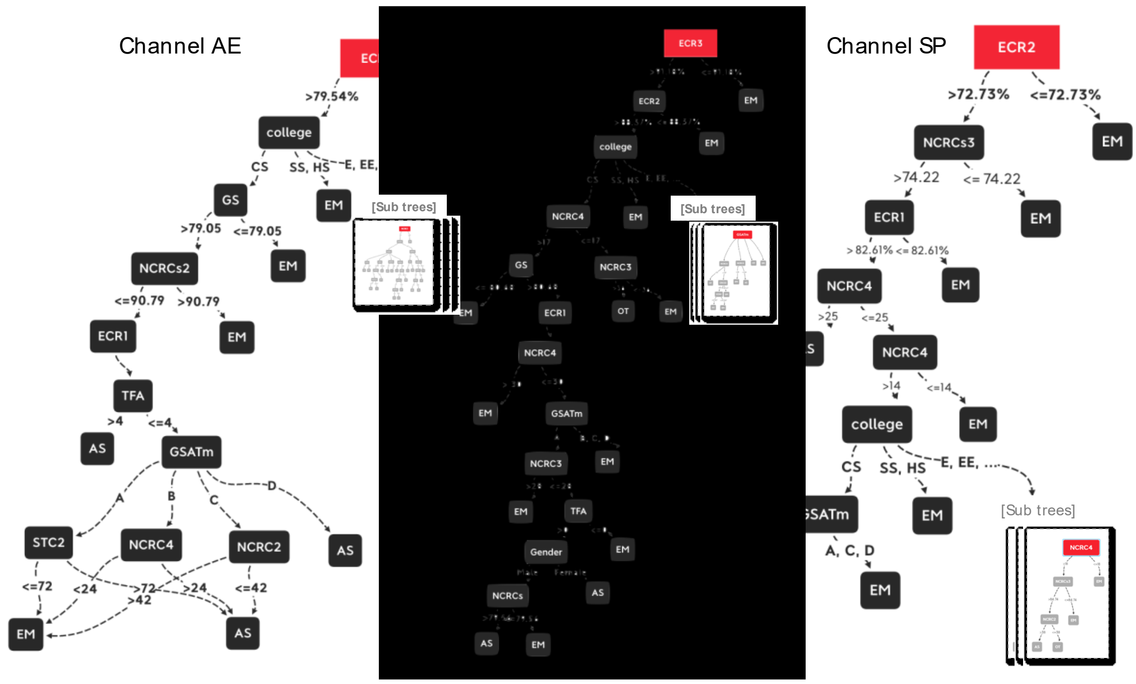


Figure 5. An example of decision tree for students recruited from three admission channel

4.4. Classification rules and information extraction from the decision table

The classification rules and information can also be represented as decision tables to facilitate decision making and promote precision education strategies. For example, Table 7 represents a decision table of classification rules to identify students who might seek to enter a master's program at their current university. It can be found that for the college of management, the GSATm, NCRC1, and NCRC4 were important features to identify whether a student had a tendency to apply to a master's program. On the other hand, NCRC2, NCRC3, NCRC4, and STC1 were useful for identifying students who were willing to enroll in a graduate institute in the current university for the college of biological science and technology. These findings suggested that it is possible to investigate the career decisions of students at different time points for different colleges. Specifically, the mathematical prior knowledge (i.e., GSATm), credits enrolled in the first year and last year before they graduate were closely related to the tendency of career decision for college of management.

Prior mathematical knowledge and credits enrolled in the first year might not significantly predict students' career decisions in the college of biological science and technology. Tracing learning performance for calculus and how many credits were enrolled in each year after the second year in university were needed to evaluate whether a student tends to engage in advanced studies in the current university. Thus, various features can be considered to investigate students' career decisions in different contexts to support help educational decision-makers and stakeholders as they evaluate precision education strategies for their institutions.

The previous sections have demonstrated a serial process of extracting information that is valuable for the practice of precision education using an institutional research perspective. For example, based on the above results, early prediction can be performed during the first year of study to allow the institution to obtain a better picture of how many students would engage in advanced studies at the university. In this way, teaching and learning resources can be evaluated for freshmen and promoted precisely in relation to how many credits students had enrolled in the previous year. To facilitate practical and interpretable applications of precision education, we presented a conceptual framework (Figure 6) for the research community in this field to explore the practice of precision education based on the comprehensive process proposed in this study.

Table 7. Examples of decision tables

ECR2	College	NCRC1	NCRC4	GSATm	STC1	NCRC3	NCRC2	CD
<= 0.795								EM
> 0.795	HC、SS							EM
	M	<=43						EM
		>43	<=21					EM
			>21	B or C				AS
				A				EM
		<=44		D				EM
		>44						AS
	BS		<=21					EM
			>21		> 67			EM
					<=67	<=34		EM
						>34	>41	AS
				> 59			<=41	AS
					<=59			EM

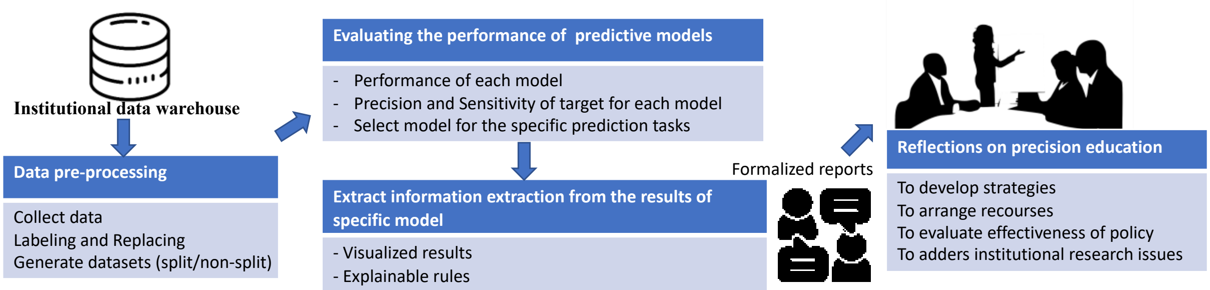


Figure 6. Conceptual framework of prediction tasks from the perspectives of institutes

5. Conclusions

Questions in institutional research for higher education have been transformed from reactive to proactive in decision and policymaking in the past decade. The detection and accommodation of individual needs play an important role in higher education because it presents an opportunity to address precision education. Among the various individual needs, career development and training are attracting attention as beneficiaries of the accurate prediction of students' career decisions.

To answer our first research question, various machine learning approaches were used as predictive models to perform predictive analyses to determine the career decisions of students with consideration for their demographics, mathematical ability, and overall learning performance. A maximum accuracy of 94.75% was achieved, and a high accuracy, kappa, precision, and sensitivity were obtained for most predictive models in this study. The results indicate that students' career decisions are predictable with reasonable accuracy.

In the pursuit of the second research question, this study sought to stimulate future research that can promote the practice of precision education. The results of this study show that the institutional database can be used to obtain precise information on the individual needs of students. This can enable specific strategies or policies to be proposed from an institutional research perspective. In this way, institutions can develop more precise interventions for their students and enable their individual needs to be met in order to reach the goal of precision education.

Overall, this study's contributions fall under three categories: theory, methodology, and application. First, it deepens the theoretical understanding of the relationships among students' demographic, mathematical knowledge, learning performance, and career decision aspects. Specifically, the predictive model identified by this study revealed that using datasets distinguished by admission channels can produce maximum performance for prediction tasks, suggesting that students recruited along different admission channels were heterogeneous. Moreover, the rates of earning credit for the second or third year and prior mathematical knowledge (i.e., GSATm) were critical features in predicting students' career decisions, regardless of their admissions channel. These findings echo previous studies, which indicated that mathematical knowledge could be related to students' career aspirations and decisions (Lazarides et al., 2020).

Second, for methodology, this study performed prediction tasks by deploying seven commonly used supervised machine learning approaches. These approaches successfully extracted relationships from students' features and career decisions. The results of this study suggested that the proposed framework (including its data structure) is appropriate for predicting students' career decisions.

Finally, a comprehensive analysis and comparison of the application of major machine learning techniques were performed in this study. It was shown that particular machine learning techniques provided for optimized prediction of different targets or purposes (e.g., specific career decisions of students in different colleges) and additional information could be obtained according to a predictive model, such as drawing decision trees or creating decision tables.

Like precision medicine, optimal precision education is a long way off (Reardon & Stuart, 2017). Emerging educational data mining techniques can be included in future analysis and comparison, such as deep learning approaches. Moreover, adequate datasets for learning data that contain a diversity of courses or learning activities and individual characteristics (e.g., motivation, interest, and preferences) are worth being made available to assemble advanced frameworks and perform prediction tasks. It would be valuable to determine whether such educational applications can help online teachers by applying the insights of precision education.

Acknowledgement

This study is supported by the Ministry of Science and Technology of the Republic of China, Taiwan, under contract numbers MOST-108-2511-H-152-015-MY2. The authors also gratefully thank Big Data Research Center (BDRC) of National Chiao-Tung University for allowing to use its data.

References

Abdullah, N., Hussin, N., Shonubi, O. A., Ghazali, S. R., & Talib, M. A. (2018). Career decision-making competence, self-knowledge, and occupational exploration: A Model for university students. *Journal of Technical Education and Training*, *10*(1), 71-81.

Adekitan, A. I., & Salau, O. (2019). The Impact of engineering students' performance in the first three years on their graduation result using educational data mining. *Heliyon*, *5*(2), e01250. doi:10.1016/j.heliyon.2019.e01250

Arthars, N., Dollinger, M., Vigentini, L., Liu, D. Y. T., Kondo, E., & King, D. M. (2019). Empowering teachers to personalize learning support. In D. Ifenthaler, D. K. Mah, & J. K. Yau (Eds.), *Utilizing learning analytics to support study success* (pp. 223-248). Cham, Switzerland: Springer.

Bahr, P. R. (2009). College hopping: Exploring the occurrence, frequency, and consequences of lateral transfer. *Community College Review, 36*(4), 271–298.

Bahr, P. R. (2010). The Bird's eye view of community colleges: A Behavioral typology of first-time students based on cluster analytic classification. *Research in Higher Education, 51*(8), 724–749.

Beemer, J., Spoon, K., Fan, J., Stronach, J., Frazee, J. P., Bohonak, A. J., & Levine, R. A. (2018). Assessing instructional modalities: Individualized treatment effects for personalized learning. *Journal of Statistics Education*, *26*(1), 31-39.

Bellgard, M. I. (2020). ERDMAS: An Exemplar-driven institutional research data management and analysis strategy. *International Journal of Information Management*, 50, 337-340.

Brunila, K. (2020). Interrupting psychological management of youth training. *Education Inquiry*, *11*(4), 302-315.

Burns, M. K., Codding, R. S., Boice, C. H., & Lukito, G. (2010). Meta-analysis of acquisition and fluency math interventions with instructional and frustration level skills: Evidence for a skill-by-treatment interaction. *School Psychology Review*, *39*(1), 69-83.

Caison, A. L. (2007). Analysis of institutionally specific retention research: A Comparison between survey and institutional database methods. *Research in Higher Education*, *48*(4), 435-451.

Camacho, D. J., & Legare, J. M. (2016). Shifting gears in the classroom—Movement toward personalized learning and competency-based education. *The Journal of Competency-Based Education*, *1*(4), 151-156.

Chen, S. Y., & Wang, J. H. (2020). Individual differences and personalized learning: A Review and appraisal. *Universal Access in the Information Society*, 1-17.

Cheng, T. M., Hou, H. Y., Agrawal, D. C., Liaw, C. F., & Chen, R. C. (2018). Remedial education for high-risk university freshmen – The Case of a university of technology in Taiwan. *Asia Pacific Journal of Education,* 38, 241-256.

Christensen, C. M., & Eyring, H. J. (2011). *The Innovative university: Changing the DNA of higher education from the inside out*. San Francisco, CA: John Wiley & Sons.

Cogan, L. S., Schmidt, W. H., & Guo, S. (2019). The Role that mathematics plays in college-and career-readiness: Evidence from PISA. *Journal of Curriculum Studies*, *51*(4), 530-553.

Collins, F. S., & Varmus, H. (2015). A New initiative on precision medicine. *New England Journal of Medicine, 372*(9), 793–795.

Collins, H., Calvo, S., Greenberg, K., Neall, L. F., & Morrison, S. (2016). Information needs in the precision medicine era: How genetics home reference can help. *Interactive Journal of Medical Research*, *5*(2), e13. doi: 10.2196/ijmr.5199

Cook, C. R., Kilgus, S. P., & Burns, M. K. (2018). Advancing the science and practice of precision education to enhance student outcomes. *Journal of School Psychology*, 66, 4-10.

Daghestani, L. F., Ibrahim, L. F., Al-Towirgi, R. S., & Salman, H. A. (2020). Adapting gamified learning systems using educational data mining techniques. *Computer Applications in Engineering Education*, *28*(3), 568-589.

Du, X., Yang, J., & Hung, J. L. (2020). An Integrated framework based on latent variational autoencoder for providing early warning of at-risk students. *IEEE Access*, *8*, 10110-10122.

Dutt, A., Ismail, M. A., & Herawan, T. (2017). A Systematic review on educational data mining. *IEEE Access*, *5*, 15991-16005.

Ellis, L. (2019). Artificial intelligence for precision education in radiology–Experiences in radiology teaching from a UK foundation doctor. *The British Journal of Radiology*, *92*(1104), 20190779. doi:10.1259/bjr.20190779

Fan, J. (2016). The Role of thinking styles in career decision-making self-efficacy among university students. *Thinking Skills and Creativity*, *20*, 63-73.

Ferrare, J. J., & Miller, J. M. (2020). Making sense of persistence in scientific purgatory: A Multi-institutional analysis of instructors in introductory science, technology, engineering, and mathematics (STEM) courses. *The Journal of Higher Education*, *91*(1), 113-138.

Flaig, M., Simonsmeier, B. A., Mayer, A. K., Rosman, T., Gorges, J., & Schneider, M. (2018). Reprint of "Conceptual change and knowledge integration as learning processes in higher education: A Latent transition analysis." *Learning and Individual Differences*, *66*, 92-104.

Freeman, V. F., Lenz, J. G., & Reardon, R. C. (2017). Career course impact on college students' career decision and affective states. *VISTAS Online*, *32*(1), 1-14.

Gulson, K. N., & Webb, P. T. (2017). Mapping an emergent field of "computational education policy": Policy rationalities, prediction and data in the age of Artificial Intelligence. *Research in Education*, *98*(1), 14-26.

Hart, S. A. (2016). Precision education initiative: Moving toward personalized education. Mind, *Brain, and Education, 10*(4), 209–211.

Healy, M., Hammer, S., & McIlveen, P. (2020). Mapping graduate employability and career development in higher education research: a citation network analysis. *Studies in Higher Education*, 1-13.

Hoffait, A. S., & Schyns, M. (2017). Early detection of university students with potential difficulties. *Decision Support Systems*, *101*, 1-11.

Kazdin, A. E., & Blase, S. L. (2011). Rebooting psychotherapy research and practice to reduce the burden of mental illness. *Perspectives on psychological science*, *6*(1), 21-37.

Ketenci, T., Leroux, A., & Renken, M. (2020). Beyond student factors: A Study of the impact on stem career attainment. *Journal for STEM Education Research*, *3*, 368-386.

Kim, E., Benson, S., & Alhaddab, T. A. (2018). A Career in academia? Determinants of academic career aspirations among PhD students in one research university in the US. *Asia Pacific Education Review*, *19*(2), 273-283.

Kokkinos, C. M., & Saripanidis, I. (2017). A Lifestyle exposure perspective of victimization through Facebook among university students. Do individual differences matter? *Computers in Human Behavior*, 74, 235-245.

Kuch, D., Kearnes, M., & Gulson, K. (2020). The Promise of precision: datafication in medicine, agriculture and education. *Policy Studies, 41*(5), 1-20.

Laanan, F. S., & Jain, D. (2016). Advancing a new critical framework for transfer student research: Implications for institutional research. *New Directions for Institutional Research*, *170*, 9-21.

Laguilles, J. S. (2016). Collecting and using postgraduate outcomes data at a private college. *New Directions for Institutional Research*, *169*, 25-36.

Lazarides, R., Dicke, A. L., Rubach, C., & Eccles, J. S. (2020). Profiles of motivational beliefs in math: Exploring their development, relations to student-perceived classroom characteristics, and impact on future career aspirations and choices. *Journal of Educational Psychology*, *112*(1), 70-92.

Lin, J. J., & Liou, P. Y. (2019). Assessing the learning achievement of students from different college entrance channels: A Linear growth curve modelling approach. *Assessment & Evaluation in Higher Education*, *44*(5), 732-747.

Lonn, S., & Koester, B. (2019). Rearchitecting data for researchers: A Collaborative model for enabling institutional learning analytics in higher education. *Journal of Learning Analytics*, *6*(2), 107-119.

Lu, O. H., Huang, A. Y., Huang, J. C., Lin, A. J., Ogata, H., & Yang, S. J. (2018). Applying learning analytics for the early prediction of Students' academic performance in blended learning. *Educational Technology & Society*, *21*(2), 220-232.

Makhluf, H. A. (2020). Precision education: Engineering learning, relevancy, mindset, and motivation in online environments. In *Exploring Online Learning Through Synchronous and Asynchronous Instructional Methods* (pp. 202-224). Hershey, PA: IGI Global.

Márquez-Vera, C., Cano, A., Romero, C., Noaman, A. Y. M., Mousa Fardoun, H., & Ventura, S. (2016). Early dropout prediction using data mining: A Case study with high school students. *Expert Systems*, *33*(1), 107-124.

Mielicki, M. K., Martinez, M. V., DiBello, L. V., Lee-Hassan, A. W., & Pellegrino, J. W. (2019). Assessing mathematics knowledge and skill: What college students actually know and can do? In O. Latkin-Troitschanskaia (Eds.), *Frontiers and Advances in Positive Learning in the Age of InformaTiOn (PLATO)* (pp. 253-265). Wiesbaden, Germany: Springer.

Miller, A. K., Osborn, D. S., Sampson Jr, J. P., Peterson, G. W., & Reardon, R. C. (2018). The Impact of a college career course on students' career decision states. *The Career Development Quarterly*, *66*(4), 371-377.

O'Connor, M. A., & Daly III, E. J. (2018). Selecting effective intervention strategies for escape-maintained academic-performance problems: Consider giving'em a break! *Journal of school psychology*, *66*, 41-53.

Palanica, A., Docktor, M. J., Lee, A., & Fossat, Y. (2019). Using mobile virtual reality to enhance medical comprehension and satisfaction in patients and their families. *Perspectives on Medical Education*, *8*(2), 123-127.

Pardo, A. (2018). A Feedback model for data-rich learning experiences. *Assessment & Evaluation in Higher Education*, *43*(3), 428-438.

Peng, Y., & Wang, Q. (2020). The Impact of mindful agency coaching and motivational interviewing on the development of positive learning dispositions in undergraduate students: A Quasi-experimental intervention study. *Journal of Educational and Psychological Consultation*, *30*(1), 63-89.

Reardon, S. F., & Stuart, E. A. (2017). Editors' introduction: Theme issue on variation in treatment effects. *Journal of Research on Educational Effectiveness, 10*(4), 671-674.

Reed, D. K., & Gates, C. (2020). Don’t fail to plan for summer reading interventions. *Preventing School Failure: Alternative Education for Children and Youth, 64*(3), 223-229.

Rienties, B., & Toetenel, L. (2016). The impact of learning design on student behaviour, satisfaction and performance: A Cross-institutional comparison across 151 modules. *Computers in Human Behavior*, *60*, 333-341.

Rodrigues, M. W., Isotani, S., & Zárate, L. E. (2018). Educational data mining: A Review of evaluation process in the e-learning. *Telematics and Informatics*, *35*(6), 1701-1717.

Ross, S. M., & Morrison, G. R. (2012). Constructing a deconstructed campus: Instructional design as vital bricks and mortar. *Journal of Computing in Higher Education*, *24*(2), 119-131.

Rump, M., Esdar, W., & Wild, E. (2017). Individual differences in the effects of academic motivation on higher education students’ intention to drop out. *European Journal of Higher Education*, *7*(4), 341-355.

Santos, A., Wang, W., & Lewis, J. (2018). Emotional intelligence and career decision-making difficulties: The Mediating role of career decision self-efficacy. *Journal of Vocational Behavior*, *107*, 295-309.

Surrette, T. N. (2020). Influence of mentoring and professional communities on the professional development of a cohort of early career secondary mathematics and science teachers. *School Science and Mathematics*, *120*(3), 175-188.

Swing, R. L., & Ross, L. E. (2016). A New vision for institutional research. *Change: The Magazine of Higher Learning*, *48*(2), 6-13.

Thompson, M. N., Her, P., Fetter, A. K., & Perez-Chavez, J. (2019). College student psychological distress: Relationship to self-esteem and career decision self-efficacy beliefs. *The Career Development Quarterly*, *67*(4), 282-297.

Ulas, O., & Yildirim, I. (2019). Influence of locus of control, perceived career barriers, negative affect, and hopelessness on career decision-making self-efficacy among Turkish university students. *International Journal for Educational and Vocational Guidance*, *19*(1), 85-109.

Valladares, M., Ramírez-Tagle, R., Muñoz, M. A., & Obregón, A. M. (2018). Individual differences in chronotypes associated with academic performance among Chilean University students. *Chronobiology international*, *35*(4), 578-583.

Warren, J. J., Clancy, T. R., Delaney, C. W., & Weaver, C. A. (2017). Big-data enabled nursing: Future possibilities. In *Big Data-Enabled Nursing* (pp. 441-463). New York, NY: Springer.

Williamson, B. (2019). Digital policy sociology: Software and science in data-intensive precision education. *Critical Studies in Education*, 1-17.

Williamson, B. J. (2020). New digital laboratories of experimental knowledge production: Artificial intelligence and education research. *London Review of Education*, *18*(2), 209-200.

Xing, W., Guo, R., Petakovic, E., & Goggins, S. (2015). Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior*, *47*, 168-181.

Yoo, J. (2019). A Study on AI education in graduate school through IPA. *Journal of The Korean Association of Information Education*, *23*(6), 675-687.

Zawacki-Richter, O., Juarros, V. M., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), 1-27. doi:10.1186/s41239-019-0171-0