Journal of
Educational Technology & Society

Published by International Forum of Educational Technology & Society
Hosted by National Taiwan Normal University, Taiwan

Volume 25 Issue 1

January 2022

ISSN: 1436-4522 (online)
ISSN: 1176-3647 (print)
http://www.j-ets.net/

Jan. 2022 Journal of Educational Technology & Society vol.25 no.1

The Journal of Educational Technology & Society has Impact Factor 3.522 and 5-Year impact factor 3.941 according to Thomson Scientific 2020 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. ET&S has achieved its purposes of 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 the Authors, Reviewers and Readers, the journal has enjoyed tremendous success.

ET&S has established a solid and stable editorial office where the Editors-in-Chief aims 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 Taiwan Normal University, Taiwan

Editorial Office
c/o Chair Professor Nian-Shing Chen, Institute for Research Excellence in Learning Sciences, Program of Learning Sciences, National Taiwan Normal University, No.162, Sec. 1, Heping E. Rd., Da-an Dist., Taipei City 10610, Taiwan.

Supporting Organizations

University of North Texas, USA
University of Piraeus, Greece
National Yunlin University of Science and Technology, Taiwan

Website Maintenance

Institute for Research Excellence in Learning Sciences, National Taiwan Normal University

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. Please place figures in the manuscript.

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 25 Number 1 2022

Table of Contents

Full Length Articles

Editorial Note

Special Issue Articles

ISSN 1436-4522 (online) and 1176-3647 (print).

Chen, C.-H., Koong, C.-S., & Liao, C. (2022). Influences of Integrating Dynamic Assessment into a Speech Recognition Learning Design to Support Students' English Speaking Skills, Learning Anxiety and Cognitive Load. *Educational Technology & Society*, *25* (1), 1-14.

Influences of Integrating Dynamic Assessment into a Speech Recognition Learning Design to Support Students' English Speaking Skills, Learning Anxiety and Cognitive Load

Chih-Hung Chen[1*], Chorng-Shiuh Koong[2] and Chien Liao[1]

[1]Master Program of Professional Teacher, National Taichung University of Education, Taiwan // [2]Department of Computer Science, National Taichung University of Education, Taiwan // duke.chchen@gmail.com // csko@mail.ntcu.edu.tw // btp107203@gm.ntcu.edu.tw

[*]Corresponding author

(Submitted January 6, 2021; Revised February 4, 2021; Accepted February 23, 2021)

ABSTRACT: Artificial intelligence (AI) technology has been progressively utilized in educational environments in recent years, due to the advances in computing and information processing techniques. The automatic speech recognition technique (ASR) provides students with instantaneous feedback and interactive oral practice for supporting a context with self-paced learning. Corrective feedback (CF) should be combined with ASR-based systems to enhance students' speaking performance, and to reduce their cognitive load. However, learners' perceptions of CF are mixed, and CF might give rise to learning anxiety. In this study, a dynamic assessment-based speech recognition (called DA-SR) learning system was designed to facilitate students' English speaking. Moreover, a quasi-experiment was implemented to evaluate the effects of the proposed approach on students' speaking learning effectiveness, via respectively providing the DA-SR and the corrective feedback-based speech recognition (called CF-SR) approaches for the experimental and control groups. The experimental results revealed that both the DA-SR group and the CF-SR group can effectively improve the students' English speaking skills, and decrease their English speaking learning anxiety. Moreover, this study further demonstrated that the DA-SR approach successfully reduced students' English class performance anxiety, and extraneous cognitive load in comparison with the CF-SR approach. It could be a valuable reference for designing English speaking learning activities in EFL learning environments.

Keywords: Artificial intelligence, Speech recognition, Corrective feedback, Dynamic assessment, Learning anxiety

1. Introduction

English is regarded as the most widely spoken language in the world. With globalization and the rapid advances in technology, English is now in widespread use, highlighting the importance of enhancing students' English competencies and global perspectives (Chen, 2020; Fu et al., 2019). Foreign language learning can be probed according to the four language skills, namely listening, reading, writing and speaking. Among these skills, enhancing speaking ability is widely deemed to be a difficult task for most English as a foreign language (EFL) learners (Gan, 2013). A great number of studies have investigated instructional approaches or contexts of EFL to promote students' English speaking skills and learning motivations (Abdullah et al., 2019). For example, Chien et al. (2020) adopted the peer assessment strategy in a spherical video-based virtual reality environment for situating learners in an authentic learning situation, and for directing them to comment on peers' English speaking performance. Such a learning strategy effectively enhanced the learners' English speaking skills and facilitated their reflections on what they had learned. Furthermore, a learner may feel anxious about public speaking or about answering questions (Bodnar et al., 2017), and accordingly reduction in learning anxiety has been regarded as a crucial factor for improving students' English speaking skills (Liu & Jackson, 2008; Zhang & Liu, 2018). Chen and Hwang (2020) asserted that speaking anxiety is related to language development, and anxiety may affect learners' oral competence. Thus, the provisions of feedback guidance and reductions in learning anxiety have been considered as significant factors for improving students' English speaking skills.

With the rapid development of information and communication technology, the ways to learn languages have changed. Language learning materials can be displayed in an interactive manner with multimedia (Hwang & Fu, 2019). Over the last few decades, computer-assisted language learning (CALL) involving diverse computer-mediated activities has attracted much attention (Fathi & Ebadi, 2020). Taking advantage of technology in an EFL class is able to facilitate practical language skills, and to reduce learning anxiety about speaking mistakes via individual practice (de Vries et al., 2015; Kuru Gönen, 2019). It is suggested that new instructional strategies or tools should be adopted to support EFL learners in promoting language skills and encouraging more interaction (Yang & Kuo, 2020). Moreover, artificial intelligence (AI) technology has been progressively utilized in educational environments in recent years, due to the advances in computing and information processing techniques. AI aims to deal with cognitive problems which are related to human intelligence, and specifically Artificial Intelligence in Education (AIED) refers to the application of AI technologies in educational contexts to facilitate instruction or decision making, such as intelligent tutoring systems and adaptive learning systems (Chen et al., 2020a; Hwang et al., 2020b). AIED can be defined from both broad and narrow perspectives, namely the use of AI techniques in education, and the utilization of machine learning (ML) or deep learning (DL) techniques in education, respectively (Chen et al., 2020b).

AIED has offered new opportunities for facilitating superior technology-enhanced learning contexts and for carrying out productive learning activities, for example, the provision of personalized learning guidance or the supply of individual needs (Chen et al., 2020a; Hwang et al., 2020b). The automatic speech recognition technique (ASR), powered by DL neural networking, refers to a kind of technology which synchronously transcribes text streams from individual speech (Shadiev et al., 2018). With the popularity of mobile devices, the adoption of ASR in EFL speaking courses is gradually increasing (Nguyen et al., 2018). ASR can provide students with instantaneous feedback and interactive oral practice for supporting a context with self-paced learning (Luo, 2016). However, ASR faces the same issue regarding cognitive overload as spoken production does. It has been asserted that CALL systems or ASR-based systems should offer automatic corrective feedback (CF) for enhancing students' speaking performance, and for reducing their cognitive load (de Vries et al., 2015; Young & Wang, 2014). Moreover, some ASR drawbacks regarding hardware and software have been described (Crescenzi-Lanna, 2020; Yang & Meinel, 2014), indicating the necessity of investigating learners' perceptions of the utilization of an ASR-based learning system.

CF in second language acquisition refers to the responses to the correctness or appropriateness of a learner's production or comprehension (Li, 2010; Li & Vuono, 2019), which is capable of providing students with both the opportunity and time for self-repaired output (de Vries et al., 2015). CF has played a vital role in the type of scaffolding that teachers offer for improving students' EFL learning (Lyster et al., 2013). Some previous research has illustrated the influences of oral CF on students' speaking skills in EFL learning, and has proposed several types of feedback to enhance students' speaking abilities (Couper, 2019). For example, Lyster and Ranta (1997) illustrated a taxonomy of six types of different corrective feedback that teachers could adopt in the classroom, namely explicit correction, recasts, clarification requests, metalinguistic feedback, elicitation, and repetition. These corrective types can be categorized into implicit and explicit feedback. Moreover, some internal and external learner variables have been proven to be significant in determining the effectiveness of CF (Penning de Vries et al., 2020). Individual learners' proficiency, motivation, and anxiety are considered internal variables, whereas learning contexts are deemed as external variables, such as CF type, outcome measures, and CALL. It is asserted that learners' perceptions of CF are mixed, and CF might give rise to learning anxiety (Bodnar et al., 2017). This implies the value of designing technology-enhanced speaking instruction, and of providing proper CF in the EFL classroom (Rassaei, 2019). Furthermore, working memory capacity is also considered to be crucial to the effects of CF, indicating the importance of probing learners' cognitive load during English speaking activities (Penning de Vries et al., 2020).

Dynamic assessment (DA), which is a kind of alternative assessment, has been referred to as a useful interactive pedagogical approach (Cho et al., 2020). One key component of DA is scaffolded feedback, which is displayed in some form of corrective feedback (Herazo et al., 2019). DA depicts learners' cognitive structures so as to diagnose learner difficulties, and to recognize potential improvements (Allal & Ducrey, 2000; Wang & Chen, 2016). For example, Antón (2009) utilized the DA approach to evaluate students' actual and emergent abilities, and facilitated the programming of individualized instruction. Furthermore, Rezaee et al. (2020) explored the potential effects of a mobile-based dynamic assessment on EFL learners' oral fluency, and verified that the students' speaking fluency was enhanced by the proposed approaches with both text-chat, and voice-chat contexts.

Collectively, in this study, a speech recognition approach with dynamic assessment was proposed. Based on the approach, a dynamic assessment-based speech recognition (called DA-SR) system was designed to facilitate students' English speaking. Furthermore, a quasi-experiment was implemented to evaluate the effects of the proposed approach on students' speaking learning effectiveness, via respectively providing the DA-SR and the corrective feedback-based speech recognition (called CF-SR) for the experimental and control groups. The research questions in this study are listed below.

- Do the students who learned with the DA-SR approach outperform those who learned with the CF-SR approach in terms of their English speaking skills?
- Do the students who learned with the DA-SR approach reveal a lower degree of learning anxiety than those who learned with the CF-SR approach?
- Can the DA-SR approach reduce the students' cognitive load in comparison with the CF-SR approach?

2. Literature review

2.1. Automatic speech recognition for English speaking skills

The automatic speech recognition technique (ASR) is deemed as a potentially valuable AI technology which can facilitate intelligible English speech of EFL students by means of immediately transcribing text streams from their speech (Huang et al., 2016; Shadiev et al., 2014). Several previous studies have emphasized the potentiality of integrating ASR with CALL for pronunciation learning (Young & Wang, 2014), such as reduced anxiety for non-native speaking (de Vries et al., 2015), positive learning motivation (Nguyen et al., 2018), and their speaking skills in the foreign language (Wang & Young, 2014). For example, Cavus and Ibrahim (2017) adopted a speech recognition technology on mobile devices for recognizing and correcting students' spoken words, and the research results revealed that the developed learning system significantly enhanced the students' English learning skills.

Information technology offers the function of repeated training, and expands the opportunities for utilizing the target language. Moreover, an individual can repeatedly conduct English speaking practices using the ASR technology, so as to improve their fluency in English (Wang & Young, 2014). ASR-based learning systems can provide students with opportunities and integrated learning stimulation for promoting their non-native oral skills via immediately evaluating English utterances (Chen, 2011). Furthermore, the integration of the ASR-based learning system affords individualized and instantaneous feedback for creating a learning context in which individual students can learn at their own pace (Luo, 2016). Hsu (2016) described that an ASR-based learning system is able to facilitate students' metacognitive strategies in language learning via offering them timely feedback, resulting in the enhancement of their pronunciation.

With the advances in mobile and wireless technology, mobile devices have great potentiality for pedagogical application in language learning (Zhang & Zou, 2020). Via the advancement of the mobile-assisted language learning systems, the significant advantages of ASR in improving EFL learners' speaking proficiency have drawn much attention (Ahn & Lee, 2016). Such a learning context is capable of reducing students' speaking anxiety for foreign English by way of providing repeated drills and self-paced learning at any time, leading to an unpressured speaking environment (Wang & Young, 2014). Moreover, students may feel anxious about speaking out in front of classmates in class. Situating them in an unpressured speaking environment using the ASR technology is capable of reducing anxiety for foreign English speaking, indicating that ASR-based CALL systems have the potential to implement excellent English speaking and conversation practice situations (de Vries et al., 2015).

Some previous research has illustrated that spoken production requires control of the articulatory system, and may lead to great cognitive load (de Vries et al., 2015). Cognitive load refers to a multidimensional construct of the cognitive system regarding the load while performing a particular task (Paas et al., 2003; Paas & van Merriënboer, 1994). Intrinsic cognitive load is considered as an inherent component of the materials itself and individual degree of prior experience, while extraneous cognitive load originates in the excess information processing caused by the instructional design (Leahy & Sweller, 2016; Wu et al., 2018). Due to the restricted working memory capacity of learners, it is crucial to explore the relation between an instructional design and cognitive load, so as to accommodate the difficulty level of the learning activities to students' learning capabilities (Hwang et al., 2020a; Lai et al., 2019). Several previous studies have asserted the significance of providing learners with automatic corrective feedback in dominating cognitive load, while adopting a CALL system (de Vries et al., 2015). Although related studies have revealed that students produced more accurate utterances with the support of corrective feedback, few have evaluated the feedback design of the ASR-based learning systems, due to lacking sufficient pedagogical approaches for the feedback provision (Young & Wang, 2014). Thus, it is critical to carefully design the presentation of corrective feedback in an ASR-based system for promoting students' speech skills, and for reducing their cognitive load.

2.2. Dynamic assessment

Dynamic assessment (DA) is one kind of alternative assessment which integrates teaching and assessment into an interactive pedagogical approach with the provision of suitable forms of mediation (Cho et al., 2020; Ebadi & Rahimi, 2019). DA aims to portray a more complete image of learners' cognitive structures for enhancing the diagnosis of students' learning difficulties and for recognizing the developmental trajectory, by means of directly measuring their replies to specific interventions (Allal & Ducrey, 2000; Wang & Chen, 2016). DA is capable of promoting learners' achievements and of probing potential abilities via offering the details of their abilities to develop the intervention programs (Swanson & Lussier, 2001). For example, Antón (2009) declared that DA empowers a deeper characterization of learners' actual and latent abilities, and advances individualized instruction that can adapt to individual needs.

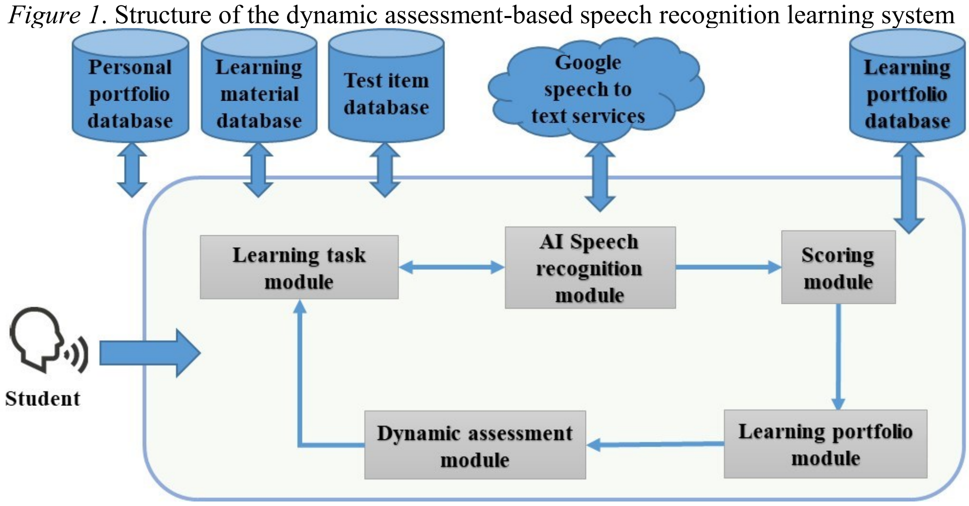
Previous research has illustrated the potential benefits of DA for improving students' learning effectiveness. Several related studies have probed the advantages of DA from the perspective of Vygotsky's zone of proximal development (ZPD), concentrating on developable abilities via intervention and interaction (Antón, 2009; Ebadi & Rahimi, 2019). For example, Wu et al. (2017) revealed that computerized dynamic adaptive tests are an effective approach for promoting learning achievement by providing individualized prompts. Bakhoda and Shabani (2019) designed a program with three sets of visual/audio/textual prompts (implicit to explicit) for evaluating emerging ZPD, and concluded that adapting to personal learning preferences with fine-tuned mediations in a computerized DA is practical. Furthermore, Rezaee et al. (2019) explored the impact of a mobile-based dynamic assessment on EFL students' oral accuracy, and declared that the proposed approach significantly improved students' oral accuracy. Andujar (2020) illustrated that DA and the dialogic mediation facilitated students' reflection on language performance, resulting in less requirement for explicit feedback and explanations.

Considering all of this evidence, it was revealed that ASR has been considered as an effective approach for enhancing students' EFL learning. However, on account of insufficient pedagogical methods for the feedback provision, few related studies have probed into the feedback mechanism designed for the ASR-based learning systems. It still remains a crucial issue to investigate the effects of integrating DA into an ASR learning context. Accordingly, this study aimed to develop a dynamic assessment-based speech recognition approach and to utilize it in an elementary school English course to evaluate its effects on students' English speaking skills, learning anxiety and cognitive load.

3. Development of the dynamic assessment-based speech recognition learning system

In this study, a dynamic assessment-based speech recognition (DA-SR) system was designed via integrating the dynamic assessment mechanism into speech recognition for enhancing students' English speech in an elementary school English course. The system was implemented utilizing PHP, MySQL, HTML, JavaScript and Google speech to text. Moreover, each student was furnished with a tablet computer and a headset for learning with no limits of time or space. Figure 1 depicts the structure of the DA-SR learning system which consists of the learning task module, the AI speech recognition module, the scoring module, the learning portfolio module, and the dynamic assessment module. Furthermore, some databases are set up to assist the modules, such as the learning material, the personal profile, the task item and the learning portfolio databases.

Figure 1. Structure of the dynamic assessment-based speech recognition learning system



The interface of the DA-SR learning system is depicted in Figure 2, which consisted of the number of the task, timing, illustration of the task, speech recognition, the prompt, and the submit button. Three kinds of learning tasks were designed in the learning activity, that is, picture reading, sentence pattern reading, and short conversations. The DA-SR learning procedure is portrayed in Figure 3. After an individual logs into the learning system, the learning task is displayed. For example, a picture of a zebra and a question, "What do you see?" are shown in the "short conversations" task. When an individual presses the "start" button and says an answer, the

DA-SR learning system immediately displays the text from the transcription of spoken language by speech recognition and requests the individual to confirm the transcription (as shown in Figure 4).

Figure 2. The interface of the DA-SR learning system

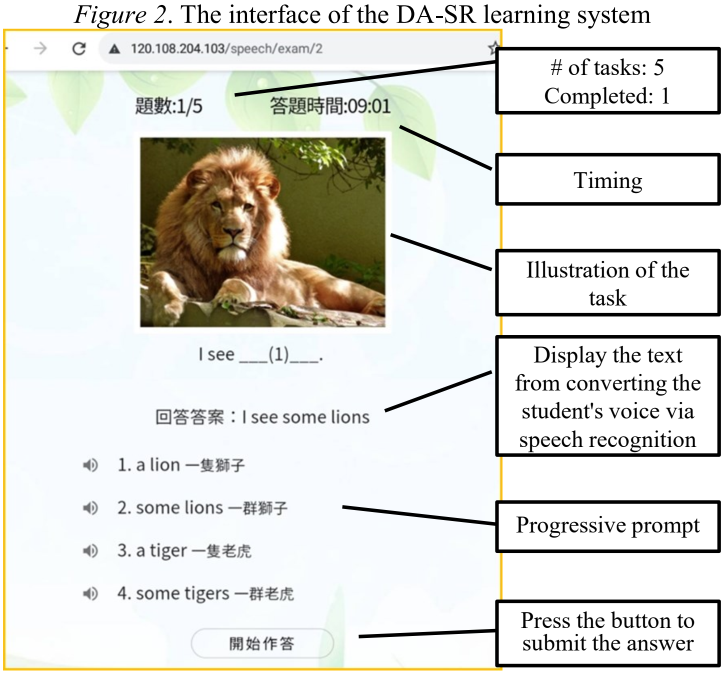
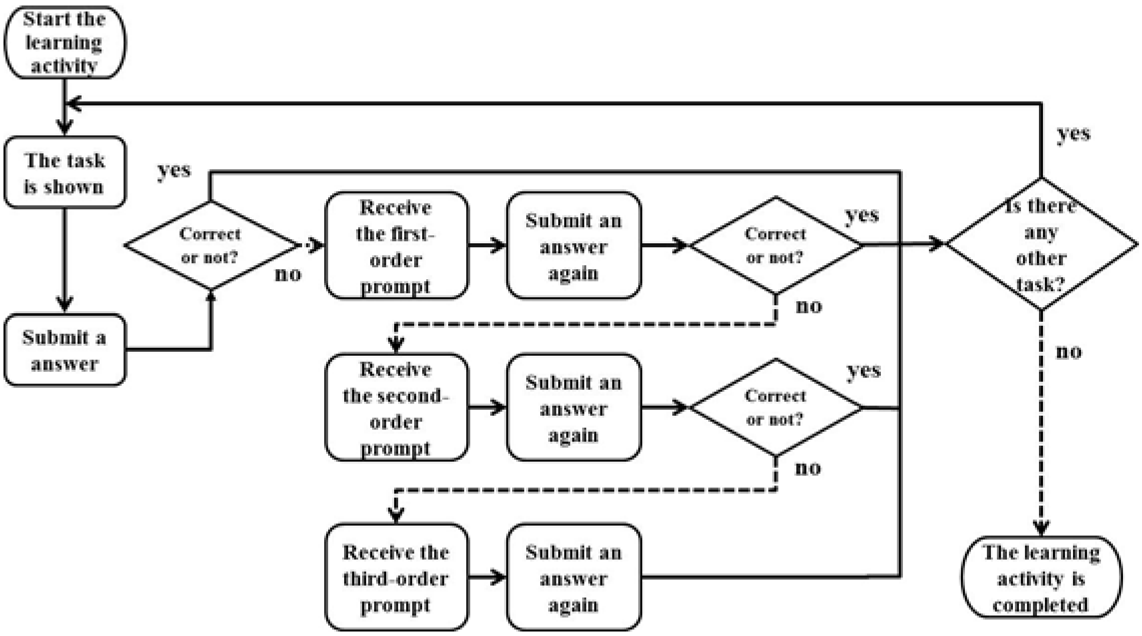


Figure 3. The learning process of the progressive prompting dynamic assessment



If the participant fails to give an appropriate answer, the DA-SR learning system assists her/him in accomplishing the task by utilizing a dynamic assessment approach. The more times a participant fails, the more concrete prompts that are given to her/him. As depicted in Figure 5, when a participant fails to produce a proper sentence the first time, the learning system provides four prompt items related to the grammar or dialogue context of the appropriate answers as the first-order prompt. If the participant fails to produce a proper sentence again, the learning system offers the Chinese translation and the application context of the four items as the second-order prompt. Furthermore, if the participant still could not submit a fitting answer the third time, the learning system then provides an audio file for demonstrating a suitable sentence.

Regarding the "picture reading" task, a picture (e.g., monkeys) is displayed on the mobile device, and the participant needs to say an answer. The three-order prompts are the provision of four prompt items (i.e., monks / monsters / monkeys / money) similar to the pronunciation of the correct answer, the supply of the Chinese translation of the prompt items, and the support of four audio files of the items in sequence. Furthermore, with respect to the "sentence pattern reading" task, a picture (e.g., a lion) and an incomplete sentence (e.g., "I see ___") are revealed on the screen. The three-order prompts are the aid of four prompt items close to the correct word or phrase (i.e., a lion / some lions / a tiger / some tigers), the support of the Chinese translation of the prompt items, and the assistance of providing four related audio files in turn. The more prompts the participants need to produce a proper sentence, the lower score they will receive. Upon successfully completing a learning task, the participant can move to the next task. When all the learning tasks are accomplished, the learning activity is completed.

Figure 4. Illustration of the transcription of spoken language

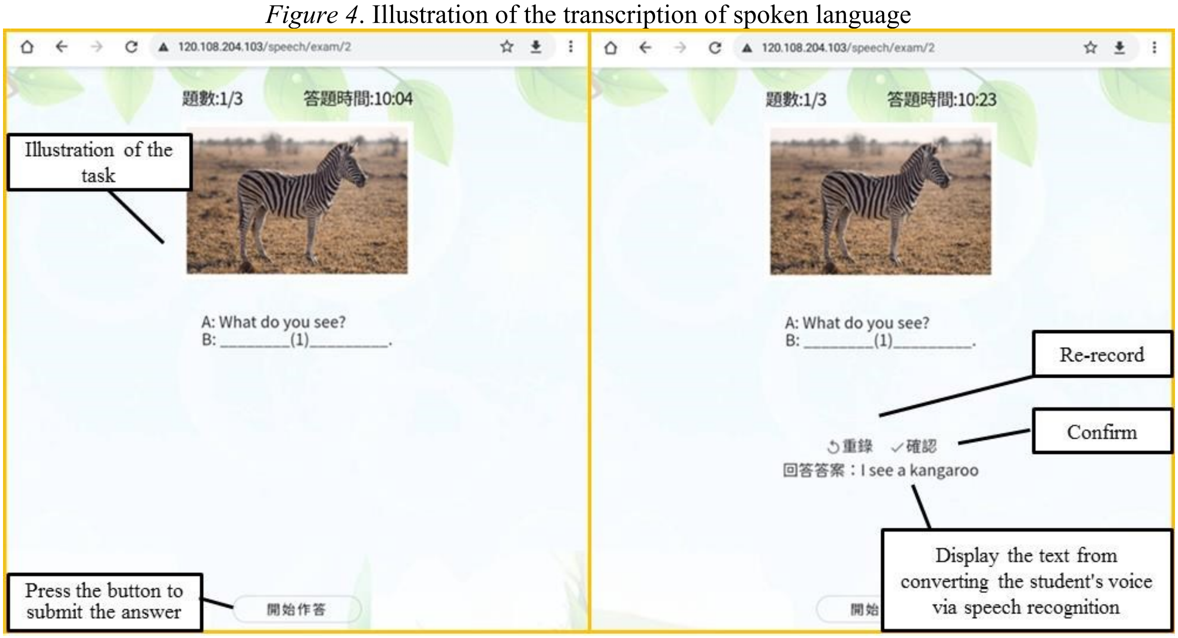
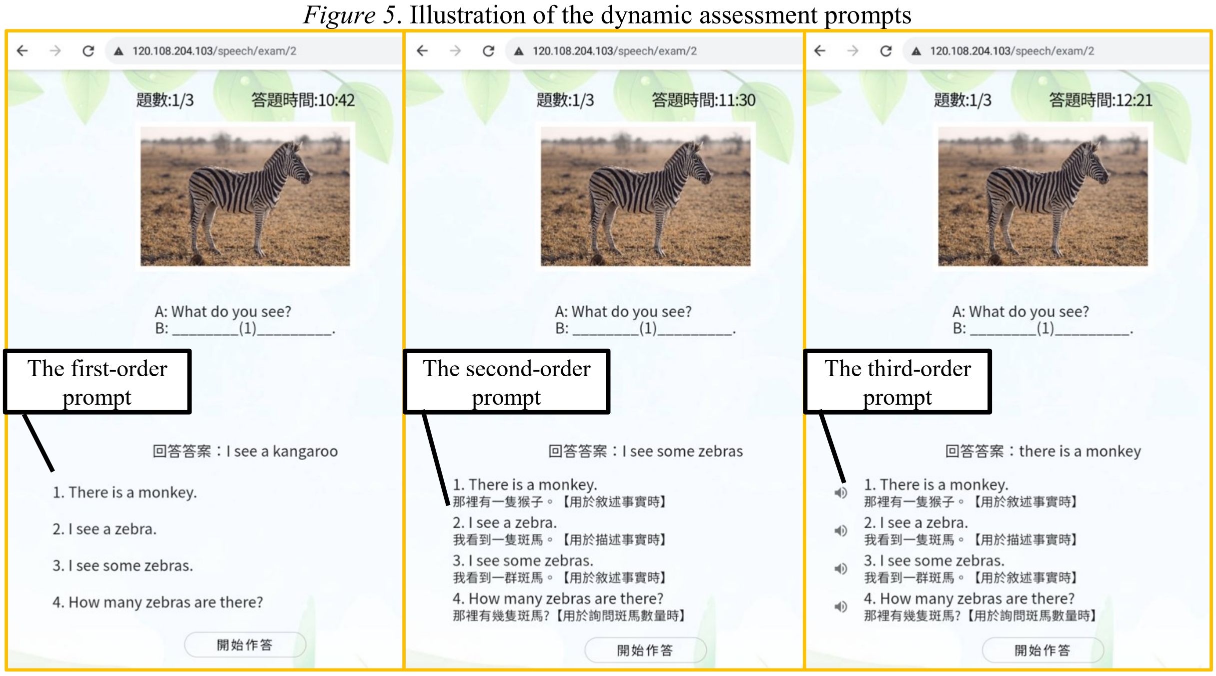


Figure 5. Illustration of the dynamic assessment prompts



4. Method

4.1. Participants

A total of 56 students from four classes of fifth graders (10- or 11-year-olds) in an elementary school in middle Taiwan were recruited for the experiment. They had English classes for three periods (a period of 40 minutes) per week. Among the four classes, two were allocated to be the experimental group (called the DA-SR group, n = 30), learning English speaking skills with the dynamic assessment-based speech recognition; the other two were the control group (called the CF-SR group, n = 26), learning English speaking skills with the corrective feedback-based speech recognition. All students in this study were already familiar with mobile technology-assisted learning.

4.2. Experimental procedure

In this study, the different English learning activities were designed to investigate the influences of integrating the dynamic assessment into a speech recognition design on the students' English speaking skills and perceptions. The experimental activity was conducted in a regular elementary school English class, and the experimental procedure is portrayed in Figure 6. First of all, the two groups experienced a regular 4-week English unit, and took the pre-test regarding English speaking skills, and filled out the pre-questionnaire about their learning anxiety.

Figure 6. The experimental procedure

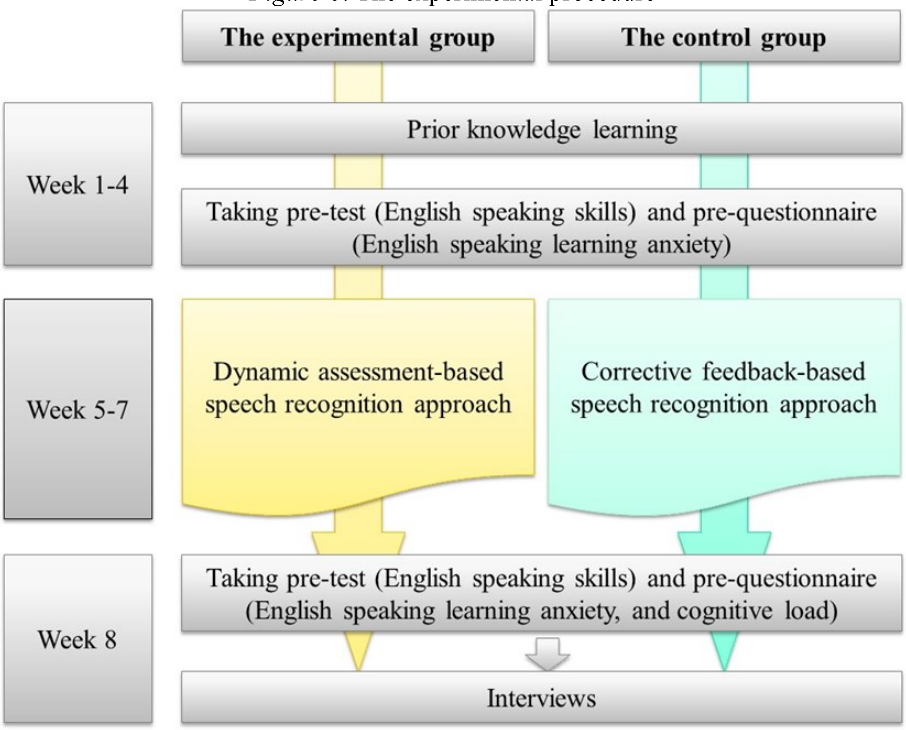


Figure 7. Snapshots of the DA-SR learning activity

Afterwards, the two groups conducted the English speaking activities with different learning approaches over 3 weeks. The learning materials, the learning tasks and the prompting content offered for the two groups were all identical, whereas the ways to display the prompts were different. When a wrong answer was selected, the learning systems activated the prompting functions. The DA-SR group was guided with the prompting content in three stages, while the CF-SR group was provided with the full prompting content all at once. Figure 7 depicts the snapshots of the learning scenarios regarding the DA-SR learning activity.

Upon completing the learning tasks, all students took the post-test concerning English speaking skills, and filled out the post-questionnaires of learning anxiety and cognitive load. Finally, a one-on-one interview was executed to investigate the views of six students recruited from each group.

4.3. Measuring tools

In this study, the pre- and post-tests of English speaking skills, and the questionnaires of learning anxiety and cognitive load were used as the instruments for assessing the students' English learning.

The pre- and post-tests were developed to evaluate the students' English speaking skills. Both tests comprised three kinds of questions with 16 items, namely short-answer questions, fill-in-the-blank items and short-answer questions about a photograph. Example items for the three kinds of questions are: "How many members are there in your family?" "_____ _____ _____ _____ in the sky" and "Are you going to the museum?" Both tests were scored on a scale of 0-80, based on the "Pre A1 Starters" assessment scale that is the first of three Cambridge English Qualifications. The assessment is formed with three criteria which are defined in candidate behaviour. The two experts who had more than 5 years' experience of teaching English courses designed the pre- and post-tests.

The learning anxiety questionnaire was modified from the instrument in Thompson and Lee's (2013) study. The original instrument consisted of four dimensions, namely "English class performance anxiety," "lack of self-confidence in English," "confidence with native speakers of English," and "fear of ambiguity in English." Furthermore, due to the learning context, the "confidence with native speakers of English" dimension and some items in the other dimensions were excluded in this study. Eventually, the adapted questionnaire of learning anxiety was composed of three scales with 18 items. The "English class performance anxiety" scale was made up of eight items (e.g., "In English class, I am so nervous that I forget what I know"). The "lack of self-confidence in English" scale included three items (e.g., "I keep considering that the other classmates speak English better than I do"). The "Fear of ambiguity in English" scale comprised seven items (e.g., "I always feel anxious about English class, although I am well prepared for it."). All items utilized a 5-point Likert rating scheme, and reverse scoring was used to re-code the responses for transforming a low point into the corresponding high point on the questionnaire. The higher the score the participants chose, the higher English learning anxiety they felt. The Cronbach's α values of the three dimensions computed by the adapted version were 0.88, 0.89, and 0.89, respectively, presenting highly acceptable reliability for rating students' English learning anxiety.

The cognitive load questionnaire was adopted from the instrument developed by Hwang et al. (2013). It had two dimensions using a 5-point Likert scale, including "mental load" and "mental effort." The mental load dimension comprised five items, while the mental effort dimension included three items. Two example items respectively for the "mental load" and "mental effort" dimensions are: "It was difficult for me to comprehend the learning content in the activity" and "It was difficult for me to follow and realize the instructional approach in the learning activity." The Cronbach's alpha coefficients of the two dimensions described by the original study were 0.85 and 0.86, respectively, showing highly acceptable reliability in internal consistency.

5. Experimental results

5.1. English speaking skills

One of the objectives of this study was to compare the impact of the DA-SR approach and that of the CF-SR approach on students' English speaking skills. Firstly, Spearman's rank correlation coefficients were computed between two sets of two experts who were recruited to judge the students' English speaking skills based on the "Pre A1 Starters" assessment scale. The Spearman's rho coefficients of the pre- and post-tests were 0.94 ($p < .01$) and 0.89 ($p < .01$), respectively, showing excellent intercoder reliability.

The paired t tests were executed to individually investigate the effects of the two learning approaches on the students' English speaking skills. Regarding the students' skills with the DA-SR approach, a significant difference was confirmed between the two tests with $t = -2.77$ ($p < .01$), as shown in Table 1. The means of the students' English speaking skills for the pre- and the post-tests respectively were 50.87 ($SD = 17.46$) and 57.50 ($SD = 17.57$). It was verified that the students who adopted the DA-SR approach significantly promoted their English speaking skills. On the other hand, with respect to the CF-SR approach, a significant difference was found between the two tests with $t = -2.71$ ($p < .05$), as depicted in Table 2. The means of the students' English speaking skills for the pre- and post-tests respectively were 41.35 ($SD = 24.24$) and 48.54 ($SD = 20.68$). It was evidenced that the students who learned with the CF-SR approach significantly improved their English speaking skills. Accordingly, both the DA-SR and the CF-SR approaches were of benefit to the students' English speaking skills.

Table 1. The paired *t*-test result of the experimental group's English speaking skills

Variable and source	*n*	Mean	*SD*	*t*
Pre-test skill	30	50.87	17.46	-2.77^{**}
Post-test skill	30	57.50	17.57	

Note. $^{**}p < .01$; Both tests were scored on a scale of 0-80.

Table 2. The paired *t*-test result of the control group's English speaking skills

Variable and source	*n*	Mean	*SD*	*t*
Pre-test skill	26	41.35	24.24	-2.71^{*}
Post-test skill	26	48.54	20.68	

Note. $^{*}p < .05$; Both tests were scored on a scale of 0-80.

Furthermore, a one-way ANCOVA was adopted to probe the influence of the different learning approaches on students' English speaking skills by excluding the interference from the two groups' prior skills. The pre-test skills were used as a covariate, while the learning approach and the post-test skills were respectively utilized as an independent variable and a dependent variable. Firstly, the homogeneity test was executed to evaluate the appropriateness of the utilization of the ANCOVA. It was proven that the assumption of homogeneity of regression was not violated with ($F = 0.01$, $p > .05$), and subsequently the ANCOVA was conducted. As shown in Table 3, no significant difference was found between the two groups' English speaking skills ($F = 0.43$, $p > .05$, $\eta^2 = 0.008$). Thus, the DA-SR approach did not benefit the students' English speaking skills in comparison with the CF-SR approach.

Table 3. The analysis of the ANCOVA on the two groups' English speaking skills

Group	*n*	Mean	*SD*	Adjusted mean	Std. error	*F*	η^2
DA-SR group	30	57.50	17.57	54.35	2.22	0.43	0.008
CF-SR group	26	48.54	20.68	52.18	2.39		

Note. Both tests were scored on a scale of 0-80.

5.2. English speaking learning anxiety

As regards the students' English speaking learning anxiety, the paired *t* tests were computed to respectively explore the impacts that the two learning approaches had on the participants. In this study, the English speaking learning anxiety included three dimensions, namely "English class performance anxiety," "lack of self-confidence in English," and "fear of ambiguity in English." As depicted in Table 4, significant differences were verified with $t = 4.98$ ($p < .001$) for the "English class performance anxiety" dimension, and $t = 5.49$ ($p < .001$) for the "fear of ambiguity" dimension, and $t = 3.72$ ($p < .01$) for the total factors in English speaking learning anxiety. This implies that the DA-SR approach can effectively reduce students' perceptions of English speaking learning anxiety, especially the "English class performance anxiety," and the "fear of ambiguity" dimensions. On the other hand, as displayed in Table 5, significant differences were found with $t = 2.49$ ($p < .05$) for the "fear of ambiguity" dimension, and $t = 2.12$ ($p < .05$) for total factors in English speaking learning anxiety, indicating that the CF-SR approach can significantly decrease students' English speaking learning anxiety, especially the "fear of ambiguity" dimension.

Moreover, the one-way ANCOVA was applied to investigate the effects of the different learning approaches on the post-questionnaire ratings of the three dimensions, and the individual pre-questionnaire ratings were utilized as the covariates. To determine whether the adoption of ANCOVA was acceptable, the homogeneity test was executed first. The homogeneity of the regression slopes was confirmed with $F= 0.13$ ($p > .05$) for the "English class performance anxiety" dimension, $F = 2.12$ ($p > .05$) for the "lack of self-confidence" dimension, and $F = 0.27$ ($p > .05$) for the "fear of ambiguity" dimension.

Following that, the ANCOVA was conducted and the results are displayed in Table 6. A significant difference was found in the post-questionnaire ratings of the students' English class performance anxiety (F = 4.08, p < .05, η^2 = 0.071), whereas no significant difference was displayed in those of their perceptions of lack of self-confidence (F = 0.83, p > .05, η^2 = 0.015), or in their perceptions of fear of ambiguity in English (F = 2.50, p > .05, η^2 = 0.045) via precluding the interference from the pre-questionnaire ratings. Furthermore, the adjusted means of the post-questionnaire ratings of the students' English class performance anxiety were 2.07 (Std. error = 0.10) for the DA-SR group, and 2.37 (Std. error = 0.11) for the CF-SR group, describing that the DA-SR approach can significantly reduce students' English class performance anxiety in comparison with the CF-SR approach. According to Cohen's (1988) assertion, the effect size for the different learning approaches was medium (η^2 > 0.059) for students' English class performance anxiety.

Table 4. The paired *t*-test result of the experimental group's English speaking learning anxiety

Factor	Variable and source	*n*	Mean	*SD*	*t*
English class performance anxiety	Pre-survey	30	2.77	1.02	4.98^{***}
	Post-survey	30	2.15	0.84	
Lack of self-confidence	Pre-survey	30	2.87	1.17	-0.13
	Post-survey	30	2.89	1.24	
Fear of ambiguity	Pre-survey	30	2.67	1.17	5.49^{***}
	Post-survey	30	2.04	0.95	
Total factors	Pre-survey	30	2.77	1.07	3.72^{**}
	Post-survey	30	2.36	0.94	

Note. $^{**}p$ < .01; $^{***}p$ < .001.

Table 5. The paired *t*-test result of the control group's English speaking learning anxiety

Factor	Variable and source	*n*	Mean	*SD*	*t*
English class performance anxiety	pre-survey	26	2.50	0.83	1.90
	post-survey	26	2.28	0.77	
Lack of self-confidence	pre-survey	26	3.12	1.30	1.29
	post-survey	26	2.78	1.20	
Fear of ambiguity	pre-survey	26	2.72	0.98	2.49^{*}
	post-survey	26	2.34	1.04	
Total factors	pre-survey	26	2.77	0.94	2.12^{*}
	post-survey	26	2.47	0.87	

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

Table 6. The ANCOVA analysis of the two groups' English speaking learning anxiety

Variable and source	Group	*n*	Mean	*SD*	Adjusted mean	Std. error	*F*	η^2
English class performance anxiety	DA-SR group	30	2.15	0.84	2.07	0.10	4.08^{*}	0.071
	CF-SR group	26	2.28	0.77	2.37	0.11		
Lack of self-confidence	DA-SR group	30	2.89	1.24	2.95	0.19	0.83	0.015
	CF-SR group	26	2.78	1.20	2.71	0.20		
Fear of ambiguity	DA-SR group	30	2.04	0.95	2.05	0.12	2.50	0.045
	CF-SR group	26	2.34	1.04	2.33	0.13		

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

5.3. Cognitive load

In this study, the cognitive load survey comprised two dimensions, namely "mental effort" and "mental load." The independent *t* tests were utilized to investigate the effects of the different learning approaches on students' intrinsic and extraneous cognitive load.

As regards the mental load dimension (as presented in Table 7), no significant difference was found in the two groups' questionnaire ratings with t = -1.99 (p > .05), describing that there is no significantly different effect of the two approaches on students' intrinsic cognitive load. On the other hand, in terms of the mental effect dimension, a significant difference was verified between the two groups' mental effort, with t = -2.17 (p < .05). The means were respectively 2.02 (SD = 0.92) and 2.60 (SD = 1.08) for the DA-SR group, and for the CF-SR group, revealing that the students who learned with the DA-SR approach were conscious of lower extraneous cognitive load than the ones who learned with the CF-SR approach. Furthermore, all the means of the two groups' questionnaire ratings were below average (Mean = 3), suggesting that all participants perceived low cognitive load during the different learning activities.

Table 7. The *t*-test result of the two groups' cognitive load levels

Variable and source	Group	*n*	Mean	*SD*	*t*
Mental load	DA-SR group	30	1.80	0.78	-1.99
	CF-SR group	26	2.32	1.14	
Mental effort	DA-SR group	30	2.02	0.92	-2.17*
	CF-SR group	26	2.60	1.08	

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

6. Discussion and conclusions

In this study, a dynamic assessment-based speech recognition approach was implemented to enhance students' English speaking learning. A learning activity was conducted in an elementary school English course. The experimental results revealed that both the experimental group (DA-SR) and the control group (CF-SR) effectively improved their English speaking skills, and decreased their perceptions of English speaking learning anxiety. Moreover, the DA-SR approach successfully reduced the students' English class performance anxiety and extraneous cognitive load in comparison with the CF-SR approach.

Speaking anxiety is regarded as a crucial factor that could affect students' speaking competence (Chen & Hwang, 2020), yet it is argued that dealing with CF could be stressful, resulting in great learning anxiety (Bodnar et al., 2017). Both the groups learning with the two different speech recognition systems significantly reduced their perceptions of English speaking learning anxiety. Such a finding corresponds to what has been depicted by Rassaei (2019), who emphasized the significance of integrating proper CF into technology-enhanced speaking instruction for EFL learning. This also confirms what has been portrayed by several reports, namely that a speech recognition approach, if properly designed, is capable of reducing learners' English speaking anxiety (de Vries et al., 2015; Wang & Young, 2014).

With respect to the three dimensions of students' English speaking learning anxiety, it is evidenced that both speech recognition approaches effectively lower students' English speaking learning anxiety for the "fear of ambiguity" dimension. This result is similar to the view asserted by Li (2010) and Li and Vuono (2019), who stated that using CF can reply to the appropriateness of a learner's production or comprehension. This could be the reason why the two speech recognition approaches with different types of CF are of great benefit in terms of reducing students' fear of ambiguity in English speaking. Moreover, only the DA-SR approach significantly decreased the students' perceptions of English class performance anxiety. It is asserted that DA is capable of adapting to individual learning preferences with fine-tuned interventions (Bakhoda & Shabani, 2019). This could explain the effect that the students who learned with the DA-SR approach revealed a lower degree of English class performance anxiety than those who learned with the CF-SR approach.

By way of illustration, Penning de Vries et al. (2020) asserted the significance of taking into consideration working memory capacity when designing CF in the English speaking learning activity. Both groups, which adopted the two speech recognition systems with the different types of CF, perceived low cognitive load. This could be a good illustration for combining CF with a speech recognition system in the English speaking learning activity. Moreover, the DA-SR approach significantly reduced the students' extraneous cognitive load in comparison with the CF-SR approach. DA and the dialogic mediation can stimulate students' reflection, and accordingly less feedback and explanations are demanded (Andujar, 2020). This could explain why the students who learned with the DA-SR approach were conscious of lower extraneous cognitive load. It is also described that excess information processing during the learning process may lead to more extraneous cognitive load (Leahy, & Sweller, 2016; Wu et al., 2018), indicating the notable value of a well-designed DA-SR approach in English speaking learning activities.

All participants in this experiment significantly enhanced their English speaking skills, implying the importance of adequately integrating CF and a speech recognition system. This accords with the view of Couper (2019) and Rassaei (2019), who described the impacts of oral CF on students' speaking abilities. As mentioned above, the CF-SR approach successfully promoted students' English speaking skills, and reduced their learning anxiety by means of properly integrating CF into a speech recognition system. This study further demonstrated that the DA-SR approach can lower students' English class performance anxiety, and extraneous cognitive load. This also supports the notion revealed by several studies (e.g., de Vries et al., 2015; Young & Wang, 2014), which emphasized that ASR-based learning systems should provide CF to promote learners' speaking skill, and to reduce their cognitive load.

This study designed the DA-SR approach for English speaking, and effectively promoted students' speaking learning effectiveness. It could be a valuable reference for designing English speaking activities in EFL learning environments. Nevertheless, neither group reduced their learning anxiety related to their self-confidence in English speaking. It is suggested that more different types of CF could be adopted in further studies regarding ASR-based learning systems. It is also worth investigating the effects of using an ASR technology in different learning contexts on students' English speaking, such as game-based learning.

Acknowledgement

This study is supported in part by the Ministry of Science and Technology of the Republic of China under Contract Numbers 109-2511-H-142-011.

References

Abdullah, M. Y., Hussin, S., & Ismail, K. (2019). Investigating the effects of the flipped classroom model on Omani EFL learners' motivation level in English speaking performance. *Education and Information Technologies, 24*(5), 2975-2995.

Ahn, T. Y., & Lee, S. M. (2016). User experience of a mobile speaking application with automatic speech recognition for EFL learning. *British Journal of Educational Technology, 47*(4), 778-786.

Allal, L., & Ducrey, G. P. (2000). Assessment of-or in-the zone of proximal development. *Learning and Instruction, 10*(2), 137-152.

Andujar, A. (2020). Mobile-mediated dynamic assessment: A New perspective for second language development. *ReCALL, 32*(2), 178-194.

Antón, M. (2009). Dynamic assessment of advanced second language learners. *Foreign Language Annals, 42*(3), 576-598.

Bakhoda, I., & Shabani, K. (2019). Bringing L2 learners' learning preferences in the mediating process through computerized dynamic assessment. *Computer Assisted Language Learning, 32*(3), 210-236.

Bodnar, S., Cucchiarini, C., Penning de Vries, B., Strik, H., & van Hout, R. (2017). Learner affect in computerised L2 oral grammar practice with corrective feedback. *Computer Assisted Language Learning, 30*(3-4), 223-246.

Cavus, N., & Ibrahim, D. (2017). Learning English using children's stories in mobile devices. *British Journal of Educational Technology, 48*(2), 625-641.

Chen, C. H. (2020). AR videos as scaffolding to foster students' learning achievements and motivation in EFL learning. *British Journal of Educational Technology, 51*(3), 657-672.

Chen, H. H. J. (2011). Developing and evaluating an oral skills training website supported by automatic speech recognition technology. *ReCALL, 23*(1), 59-78.

Chen, M. R. A., & Hwang, G. J. (2020). Effects of a concept mapping-based flipped learning approach on EFL students' English speaking performance, critical thinking awareness and speaking anxiety. *British Journal of Educational Technology, 51*(3), 817-834.

Chen, X., Xie, H., & Hwang, G. J. (2020a). A Multi-perspective study on Artificial Intelligence in Education: Grants, conferences, journals, software tools, institutions, and researchers. *Computers and Education: Artificial Intelligence, 1,* 100005. https://doi.org/10.1016/j.caeai.2020.100005

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. https://doi.org/10.1016/j.caeai.2020.100002

Chien, S. Y., Hwang, G. J., & Jong, M. S. Y. (2020). Effects of peer assessment within the context of spherical video-based virtual reality on EFL students' English-Speaking performance and learning perceptions. *Computers & Education, 146,* 103751. https://doi.org/10.1016/j.compedu.2019.103751

Cho, E., Compton, D. L., & Josol, C. K. (2020). Dynamic assessment as a screening tool for early identification of reading disabilities: A Latent change score approach. *Reading and Writing, 33*(3), 719-739.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Couper, G. (2019). Teachers' cognitions of corrective feedback on pronunciation: Their beliefs, perceptions and practices. *System, 84*, 41-52.

Crescenzi-Lanna, L. (2020). Multimodal learning analytics research with young children: A Systematic review. *British Journal of Educational Technology, 51*(5), 1485-1504.

de Vries, B. P., Cucchiarini, C., Bodnar, S., Strik, H., & van Hout, R. (2015). Spoken grammar practice and feedback in an ASR-based CALL system. *Computer Assisted Language Learning, 28*(6), 550-576.

Ebadi, S., & Rahimi, M. (2019). Mediating EFL learners' academic writing skills in online dynamic assessment using Google Docs. *Computer Assisted Language Learning, 32*(5-6), 527-555.

Fathi, J., & Ebadi, S. (2020). Exploring EFL pre-service teachers' adoption of technology in a CALL program: Obstacles, motivators, and maintenance. *Education and Information Technologies, 25*, 3897-3917.

Fu, Q. K., Lin, C. J., Hwang, G. J., & Zhang, L. (2019). Impacts of a mind mapping-based contextual gaming approach on EFL students' writing performance, learning perceptions and generative uses in an English course. *Computers & Education, 137*, 59-77.

Gan, Z. (2013). Understanding English speaking difficulties: An Investigation of two Chinese populations. *Journal of Multilingual and Multicultural Development, 34*(3), 231-248.

Herazo, J. D., Davin, K. J., & Sagre, A. (2019). L2 dynamic assessment: An Activity theory perspective. *The Modern Language Journal, 103*(2), 443-458.

Hsu, L. (2016). An Empirical examination of EFL learners' perceptual learning styles and acceptance of ASR-based computer-assisted pronunciation training. *Computer Assisted Language Learning, 29*(5), 881-900.

Huang, Y. M., Shadiev, R., & Hwang, W. Y. (2016). Investigating the effectiveness of speech-to-text recognition applications on learning performance and cognitive load. *Computers & Education, 101*, 15-28.

Hwang, G. J., & Fu, Q. K. (2019). Trends in the research design and application of mobile language learning: A Review of 2007–2016 publications in selected SSCI journals. *Interactive Learning Environments, 27*(4), 567-581.

Hwang, G. J., Sung, H. Y., Chang, S. C., & Huang, X. C. (2020a). 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*, 100003. https://doi.org/10.1016/j.caeai.2020.100003

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020b). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence, 1*, 100001. https://doi.org/10.1016/j.caeai.2020.100001

Hwang, G. J., Yang, L. H., & Wang, S. Y. (2013). A Concept map-embedded educational computer game for improving students' learning performance in natural science courses. *Computers & Education, 69*, 121-130.

Kuru Gönen, S. İ. (2019). A Qualitative study on a situated experience of technology integration: Reflections from pre-service teachers and students. *Computer Assisted Language Learning, 32*(3), 163-189.

Lai, A. F., Chen, C. H., & Lee, G. Y. (2019). An Augmented reality-based learning approach to enhancing students' science reading performances from the perspective of the cognitive load theory. *British Journal of Educational Technology, 50*(1), 232-247.

Leahy, W., & Sweller, J. (2016). Cognitive load theory and the effects of transient information on the modality effect. *Instructional Science, 44*(1), 107-123.

Li, S. (2010). The Effectiveness of corrective feedback in SLA: A Meta-analysis. *Language Learning, 60*(2), 309-365.

Li, S., & Vuono, A. (2019). Twenty-five years of research on oral and written corrective feedback in System. *System, 84*, 93-109.

Liu, M., & Jackson, J. (2008). An Exploration of Chinese EFL learners' unwillingness to communicate and foreign language anxiety. *The Modern Language Journal, 92*(1), 71-86.

Luo, B. (2016). Evaluating a computer-assisted pronunciation training (CAPT) technique for efficient classroom instruction. *Computer Assisted Language Learning, 29*(3), 451-476.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition, 20*, 37-66.

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching, 46*(1), 1-40.

Nguyen, T. H., Hwang, W. Y., Pham, X. L., & Ma, Z. H. (2018). User-oriented EFL speaking through application and exercise: Instant speech translation and shadowing in authentic context. *Educational Technology & Society, 21*(4), 129-142.

Paas, F. G. W. C, Tuovinen, J. E., Tabbers, H., & van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63-71.

Paas, F., & van Merriënboer, J. J. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*(4), 351-371.

Penning de Vries, B. W., Cucchiarini, C., Strik, H., & Van Hout, R. (2020). Spoken grammar practice in CALL: The Effect of corrective feedback and education level in adult L2 learning. *Language Teaching Research, 24*(5), 714-735.

Rassaei, E. (2019). Computer-mediated text-based and audio-based corrective feedback, perceptual style and L2 development. *System, 82*, 97-110.

Rezaee, A. A., Alavi, S. M., & Razzaghifard, P. (2019). The Impact of mobile-based dynamic assessment on improving EFL oral accuracy. *Education and Information Technologies, 24*(5), 3091-3105.

Rezaee, A. A., Alavi, S. M., & Razzaghifard, P. (2020). Mobile-based dynamic assessment and the development of EFL students' oral fluency. *International Journal of Mobile Learning and Organisation, 14*(4), 511-532.

Shadiev, R., Hwang, W. Y., Chen, N. S., & Huang, Y. M. (2014). Review of speech-to-text recognition technology for enhancing learning. *Educational Technology & Society, 17*(4), 65-84.

Shadiev, R., Wu, T. T., Sun, A., & Huang, Y. M. (2018). Applications of speech-to-text recognition and computer-aided translation for facilitating cross-cultural learning through a learning activity: Issues and their solutions. *Educational Technology Research and Development, 66*(1), 191-214.

Swanson, H. L., & Lussier, C. M. (2001). A Selective synthesis of the experimental literature on dynamic assessment. *Review of Educational Research, 71*(2), 321-363.

Thompson, A. S., & Lee, J. (2013). Anxiety and EFL: Does multilingualism matter? *International Journal of Bilingual Education and Bilingualism, 16*(6), 730-749.

Wang, J. R., & Chen, S. F. (2016). Development and validation of an online dynamic assessment for raising students' comprehension of science text. *International Journal of Science and Mathematics Education, 14*(3), 373-389.

Wang, Y. H., & Young, S. S. C. (2014). A Study of the design and implementation of the ASR-based iCASL system with corrective feedback to facilitate English learning. *Educational Technology & Society, 17*(2), 219-233.

Wu, H. M., Kuo, B. C., & Wang, S. C. (2017). Computerized dynamic adaptive tests with immediately individualized feedback for primary school mathematics learning. *Educational Technology & Society, 20*(1), 61-72.

Wu, P. H., Hwang, G. J., Yang, M. L., & Chen, C. H. (2018). Impacts of integrating the repertory grid into an augmented reality-based learning design on students' learning achievements, cognitive load and degree of satisfaction. *Interactive Learning Environments, 26*(2), 221-234.

Yang, H., & Meinel, C. (2014). Content based lecture video retrieval using speech and video text information. *IEEE Transactions on Learning Technologies, 7*(2), 142-154.

Yang, Y. F., & Kuo, N. C. (2020). New teaching strategies from student teachers' pedagogical conceptual change in CALL. *System, 90*, 102218. https://doi.org/10.1016/j.system.2020.102218

Young, S. S. C., & Wang, Y. H. (2014). The Game embedded CALL system to facilitate English vocabulary acquisition and pronunciation. *Educational Technology & Society, 17*(3), 239-251.

Zhang, R., & Zou, D. (2020). Influential factors of working adults' perceptions of mobile-assisted vocabulary learning with multimedia annotations. *International Journal of Mobile Learning and Organisation, 14*(4), 533-548.

Zhang, Y., & Liu, L. (2018). Using computer speech recognition technology to evaluate spoken English. *Educational Sciences: Theory & Practice, 18*(5), 341-1350.

Chatbot-facilitated Nursing Education: Incorporating a Knowledge-Based Chatbot System into a Nursing Training Program

Ching-Yi Chang[1,2], Shu-Yu Kuo[1] and Gwo-Haur Hwang[3*]

[1]School of Nursing, College of Nursing, Taipei Medical University, Taiwan // [2]Department of Nursing, Shuang Ho Hospital, Taipei Medical University, New Taipei City, Taiwan, Republic of China // [3]Bachelor Program in Industrial Technology, National Yunlin University of Science and Technology, Taiwan // frinng.cyc@gmail.com // sykuo@tmu.edu.tw // ghhwang0424@gmail.com

[*]Corresponding author

(Submitted January 11, 2021; Revised February 11, 2021; Accepted March 13, 2021)

ABSTRACT: Conventional nursing courses have solely adopted lecture-based instruction for knowledge delivery, which tends to lack interaction, rehearsal, and personalized feedback. The development of chatbot technologies and their broad application have provided an opportunity to solve the abovementioned problems. Some knowledge-based chatbot systems have been developed; however, it is still a challenging issue for researchers to determine exactly how to effectively apply these chatbot technologies in nursing training courses. Intending to explore the application mode of chatbot technologies and their effectiveness in nursing education, this study integrated a knowledge-based chatbot system into the teaching activities of a physical examination course, using smartphones as the learning devices, and guiding students to practice their anatomy knowledge in addition to analyzing their learning efficacy and pleasure. A quasi-experiment was conducted by recruiting two classes of university students with nursing majors. One class was the experimental group learning with the knowledge-based chatbot system, while the other class was the control group learning with the traditional instruction. Based on the experimental results, the knowledge-based chatbot system effectively enhanced students' academic performance, critical thinking, and learning satisfaction. The results indicate that the application of chatbots has great potential in nursing education.

Keywords: Chatbots, Knowledge-based chatbot system, Nursing Training, Mobile learning

1. Introduction

Physical examination is the most common method used for the diagnosis of diseases, and it is the basic foundation of understanding the physiological structures and their characteristics. Human Anatomy refers to the program that systematically introduces each of the human organ systems, such as the nervous system and muscle distribution (Kurniawan & Witjaksono, 2018). It aims to cultivate students' concepts of human organ systems from both macro and micro perspectives, and enable them to have the ability to combine physical examination in the workplace, that is, performing accurate inspection, auscultation, palpation, and percussion. As such, they will be able to gather the physical examination data of inpatients and give it to doctors for further assistance and accurate diagnosis of diseases (Narula et al., 2018). Intending to offer students a sense of authenticity, the traditional training in Medical Schools has been supported by voluntary body donation to assist teachers in guiding medical students to have a deep understanding of the structures or systems of different organs, nerves, and muscle distribution in the human body.

Nowadays, challenges have been posed to medical study because of the rapid changes in and advancement of medical knowledge owing to new diseases and medicines (Innocent, 2016; Rather et al., 2017). With the aim of helping students to make correct decisions when dealing with real cases, it is necessary to engage them in authentic problem-solving contexts (Hwang & Chang, 2020; Trasmundi & Linell, 2017). Accordingly, the notation of "precision medicine" has been proposed; this refers to the process of making precise medical decisions based on detailed and well-analyzed information (Kim, 2019). Cook et al. (2018) referenced the analysis of empirical evidence in precision education and formulated the best strategy to intervene in some students' learning so as to enhance their learning achievements including diagnosis, prediction, counseling, and prevention. One of the basic competences for conducting prevision medicine is physical assessment. No matter serving in which types of medical institutions, medical staffs are requested to learn and pass the exam of physical assessment before entering the workplace.

Researchers have further indicated that one potential solution to this aim is to apply chatbot technologies with the provision of personalized learning guidance and feedback in nursing training programs (Chang et al., 2021; Hernandez, 2019; Tsai et al., 2020). It is a possible direction which fits the goals of precision education (Tsai et al., 2020). Several previous studies have further reported the importance of guiding students to think in depth (e.g., why this happens and how to deal with a problem?) and to search for additional information to complete an assessment-oriented learning tasks; for example, Chu et al. (2010) developed employed a two-tier test approach in a mobile learning context of a natural science course; Chou et al. (2007) also used the two-tier test to guide students to think in depth in the learning activities related to digital copyright laws; Hwang and Chang (2011) adopted a formative assessment approach in a mobile learning activity of an elementary school social studies course. Although such assessment-based approaches have been recognized to be effective, the students were generally "guided to think" rather than "encouraged to explore and make decisions." To this end, many scholars have attempted to apply chatbot technologies to boost students' active learning behaviors as well as enhancing their learning efficacy (Tegos & Demetriadis, 2017). For instance, Beaumont (1994) applied a human anatomy tutorial system in the medical courses in a university, while Tegos et al. (2016) applied a chatbot-based agent which permitted learners to practice making decisions in a virtual learning environment, and found the approach effective in terms of improving students' learning performance in an "academically productive talk" course.

As mentioned in the above literature, educational technology researchers have started examining how the usage of chatbot technologies can boost the efficiency of teaching and learning. Simultaneously, learners can connect the learned knowledge with the actual problems encountered in their practice through the use of chatbot systems. For most nursing students, physical examination, which is a complex procedure aiming at making judgment on a patient's physical status based on the data collected by observing and inquiring the patient as well as the medical test results, has been recognized as an important and challenging task. The nursing students need to have accurate physiological knowledge as well as the competences to execute inspection, auscultation, percussion, and palpation to correctly complete a physical examination task (Narula et al., 2018). With the aim of solving these problems, this research attempted to integrate a knowledge-based chatbot system into a physical examination course and to overcome the shortcomings of traditional teaching in order to improve students' learning efficacy. A knowledge-based chatbot system refers to the chatbot that communicates with users based on the expertise stored in a knowledge base (Kumar, 2020). Several previous studies have reported that the use of knowledge-based systems to provide learning supports or guidance in making decisions has great potential in improve students' critical thinking (Jerome et al., 2019) and learning satisfaction (Chen, 2012). Intending to prove the effectiveness of this model, this experiment aimed to answer the following questions:

- Does the integration of the knowledge-based chatbot system into the learning mode of physical examination facilitate students' learning achievement when compared to conventional teaching?
- Does the integration of the knowledge-based chatbot system into the learning mode of physical examination improve students' critical thinking when compared to conventional teaching?
- Does the integration of the knowledge-based chatbot system into the learning mode of physical examination effectively improve students' learning satisfaction when compared to conventional teaching?

2. Literature review

The term chatbot refers to a computer application or system which interacts with users in a chat-based mode using natural language (Hwang & Chang, 2021; Smutny & Schreiberova, 2020). There are several roles of chatbots such as information or knowledge providers (Lan, 2020), convention partners (Shawar, 2017), interactive agents (Erickson & Kim, 2020), learning partners (Fryer et al., 2019) and tutors (Pérez et al., 2020). For example, when users ask some questions or raise new topics, chatbots respond with natural language-like statements based on the data or knowledge stored in the database (Balsmeier, 2018; Smutny & Schreiberova, 2020). Researchers have indicated that such a natural language-based interactive mode makes chatbots highly accepted by most people (Chang & Tseng, 2019; Shorey et al., 2019). Stuij et al. (2020) further stated that the use of chatbots could improve learners' communication skills.

In the past decade, researchers have applied chatbots to several application domains, including learning about employability issues (Wang et al., 2021; Ward et al., 2016), social networks (Pérez-Soler et al., 2018), specific languages (Pérez-Soler et al., 2019), learning Chinese (Chen et al., 2020a), basic computer learning (Yin et al., 2020), and healthcare and smart home domains (Valtolina et al., 2020). For example, Samarakou et al. (2018) found that the usage of chatbots in an informatics course in a university can provide learning opportunities for diagnosis, guidance, and assessment, which resulted in improvements in the students' learning efficacy. Lin and Chang (2020) applied a chatbot in a post-secondary writing course and found that the approach improved the students' writing quality more than traditional instruction did. It can be predicted that advances in wireless networks, sensing technology, and mobile technology will further encourage the use of chatbots in various applications, as indicated by Yin et al. (2020).

Researchers have also tried to employ chatbots in different ways based on the educational objectives and contexts (Abbasi & Kazi, 2014; Van Seters et al., 2012). For instance, Xin et al. (2020) proposed a conceptual model to train students to solve problems with learned knowledge, through the means of analyzing the subjective materials and conducting tests with the provision of learning suggestions, aiming at better assistance for them to combine the knowledge learned from textbooks. Thirumalaraju et al. (2019) proposed the installation of an application on smartphones, and suggested applying chatbot technologies to online health education, in particular enabling patients to receive education on personal hygiene and personal healthcare, as well as enabling them to make relevant decisions to manage their health goals. The implementation of online education for healthcare and obesity management in the United States is an example that illustrates this idea. Moreover, Song et al. (2019) suggested an interaction between online courses, chatbot evaluating strategies, and relevant academic content stated in the literature. The course coordinators can flexibly maintain the content of academic courses, conduct virtual conferences, and provide announcements. The results have shown that participants agree unanimously with the benefits of applying chatbots to online courses.

In addition to the user interface, scholars have emphasized the key to the value of chatbots, that is, the information or knowledge included in the chatbot systems (Shum et al., 2018; Smutny & Schreiberova, 2020; Tegos & Demetriadis, 2017). For example, Beattie, Edwards, and Edwards (2020) indicated that the positive impacts of chatbots in education highly rely on the quality and quantity of the information and knowledge included. This implies the importance of incorporating an effective knowledge or data collection mechanism in chatbots (Sheth et al., 2019). A knowledge-based chatbot is a chatbot system that includes a mechanism to extend the quality content in the database (Kapočiūtė-Dzikienė, 2020). Knowledge-based systems emphasized the use of knowledge provided by domain experts to solve problems (Zhang et al., 2017). Researchers have indicated that the knowledge base is the key to enable computer systems to imitate intelligent behaviors of human (Hwang et al., 2020a; Yulianto et al., 2020). Some researchers have predicted that chatbots can even play the role of "smart teachers," "smart learning partners" or "smart students" in educational settings if domain knowledge can be properly acquired, organized and employed in chatbots using knowledge acquis ion or machine learning approaches (e.g., deep learning) (Darshan Singh et al., 2018). For example, Smutny and Schreiberova (2020) reported the trends of using chatbots to analyze individual students' learning status and provide personalized learning paths, user interfaces and learning content. The advancement of wireless communication and sensory technology has further provided an environment for applying chatbots in diverse ways, and has led to the innovative thinking of educational researchers in implementing chatbots in education studies, such as guiding students to solve problems in the real-life environment with the support of chatbot applications (Chang & Hwang, 2018). As a result, the use of chatbot technologies has gradually changed the role of teachers in school settings. Teachers therefore have more time to guide students to think, practice and apply knowledge based on individual students' needs. This can assist teachers in improving the quality of their teaching (Hsu, 2020).

It can be seen that education has become more humane and personalized, which can enhance students' learning achievement (Chang et al., 2018). There is, therefore, an increasing need to consider individual differences in developing digital learning systems and to analyze the applications of chatbots in education (Yin et al., 2020). Educational developers have not only engaged in innovative research and teaching, but have also adopted technologies to help students learn efficiently in professional training without being limited by location or time through an integration of cross-field cooperation (Tsai et al., 2020). This research, therefore, applied a knowledge-based chatbot system to improve university students' learning efficacy in a physical examination course.

3. Experimental design

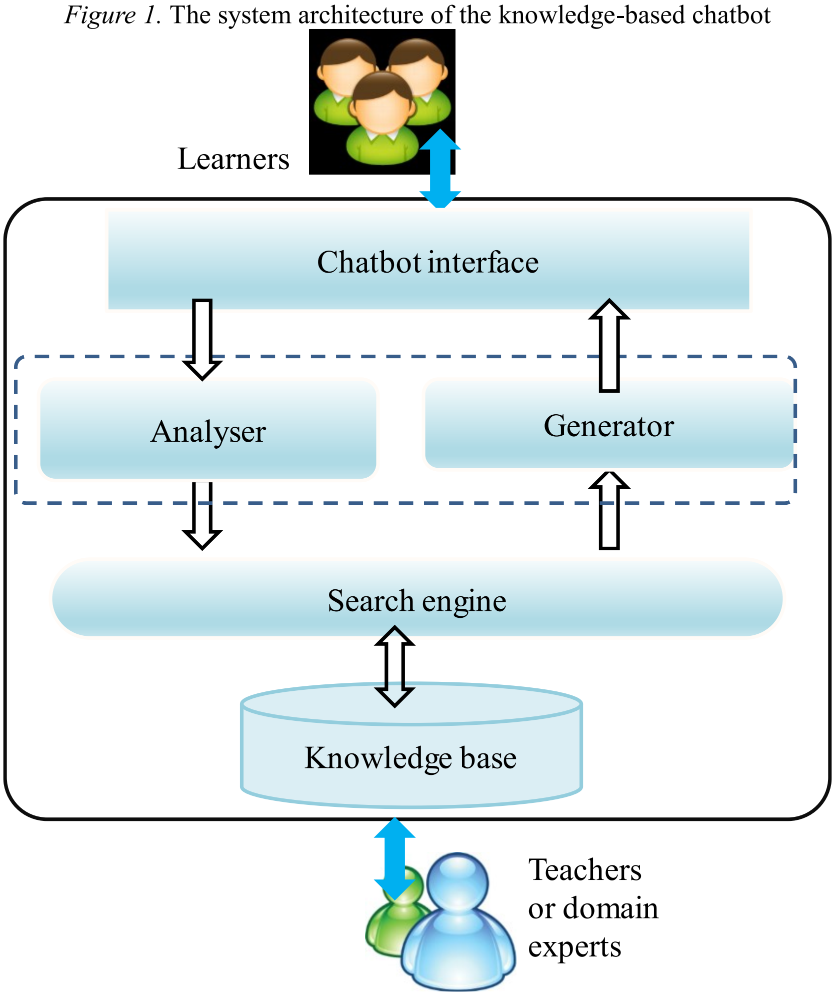
3.1. Participants

This research intended to show the effectiveness of a knowledge-based chatbot system in a nursing school in Taiwan by allowing students to attend training on physical examination, which is a compulsory course for the basic care in nursing schools and hospitals, and a necessary component of nursing training. Intending to evaluate the effectiveness of the proposed method, this study included an experimental group and a control group. A total of 32 nursing students with an average age of 21 participated in this study, with 16 students in the experimental group learning with the knowledge-based chatbot system, and the other 16 students in the control group learning via conventional teaching of physical examination. In order to compare the academic performance, critical thinking and learning satisfaction of the students in the two groups, they were asked to take a pre-test and complete a pre-questionnaire.

3.2. The knowledge-based chatbot

A knowledge-based system refers to the systems that make decisions or provide assistance based on the expertise stored in a knowledge base (Chen et al., 2020b; Hwang et al., 2020; Saura et al., 2019). The expertise in the knowledge base could be domain knowledge as well as experts' experiences of making decisions on different cases (Abbas et al., 2021; Gil et al., 2019; Hwang et al., 2020b). With the aim of educating students about Human Anatomy, this research adopted a knowledge-based system named "Anatomy Quiz," which was developed by Alexander Streuer (see https://is.gd/b6j77n). "Anatomy Quiz" uses the concepts of the tree searching algorithm and rational agent to establish a medical knowledge database. This knowledge-based system has 56 courses and 833 anatomical structures including anatomical knowledge of the skeleton, muscles, and organs. As the original "Anatomy Quiz" system is a knowledge based system, a chatbot interface was provided in the present study to enable the students to use the knowledge base in an interactive way, as shown in Figure 1. When a student talks to the chatbot, the "Analyzer" interprets the sentences submitted by the student, and searches for the relevant information from the knowledge base. The "Generator" then summarizes and organizes the searched information, and replies to the student.

Figure 1. The system architecture of the knowledge-based chatbot



This knowledge-based chatbot provides three different interactive learning models: (1) selecting and marking anatomical structures; (2) providing correct professional vocabulary such as *humerus*; and (3) making a diagnosis of the tagged anatomical structure, as shown in Figure 2.

Students can learn professional vocabulary from each system and make a diagnosis of the tagged anatomical structure through "Anatomy Quiz," thus increasing the interactivity of their learning. Moreover, they can repeatedly take the quiz to familiarize themselves with the knowledge of anatomical structures such as bones, muscles, and organs. During the learning process, the knowledge-based chatbot system guides the student in the learning tasks related to the selected topic, and in answering a series of questions. If the student makes a correct decision or choice in the specified time, the record is updated, as shown in Figure 3; otherwise, the system provides hints or complementary materials as well as calculating the time the student spent on the task and updating the record.

Figure 4 shows the learning scenario of using the knowledge-based chatbot system to learn via smartphones. One of the learning tasks was to determine the type of heart conditions based on the patient's' systolic heart murmurs. The students can interact with the chatbot to find evidences to support their assumptions.

Figure 2. User interface of the knowledge-based chatbot system

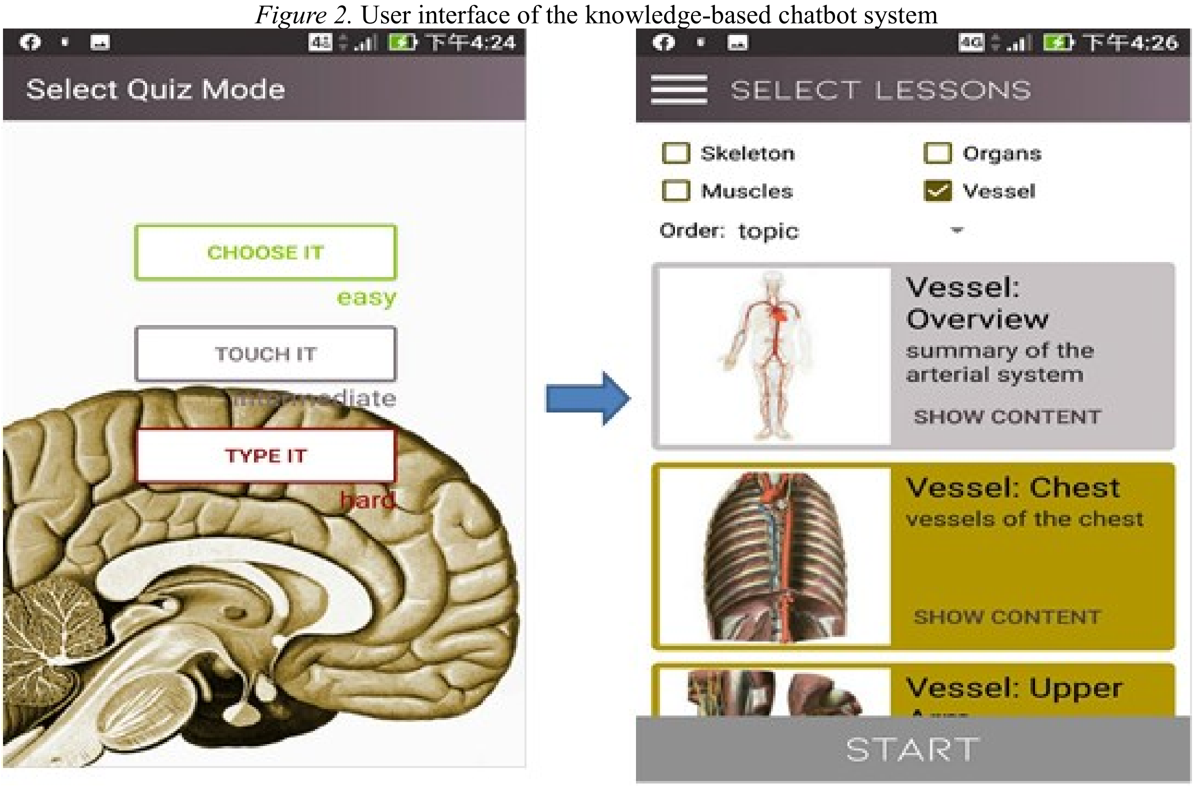


Figure 3. Interactive screen of the knowledge-based chatbot system

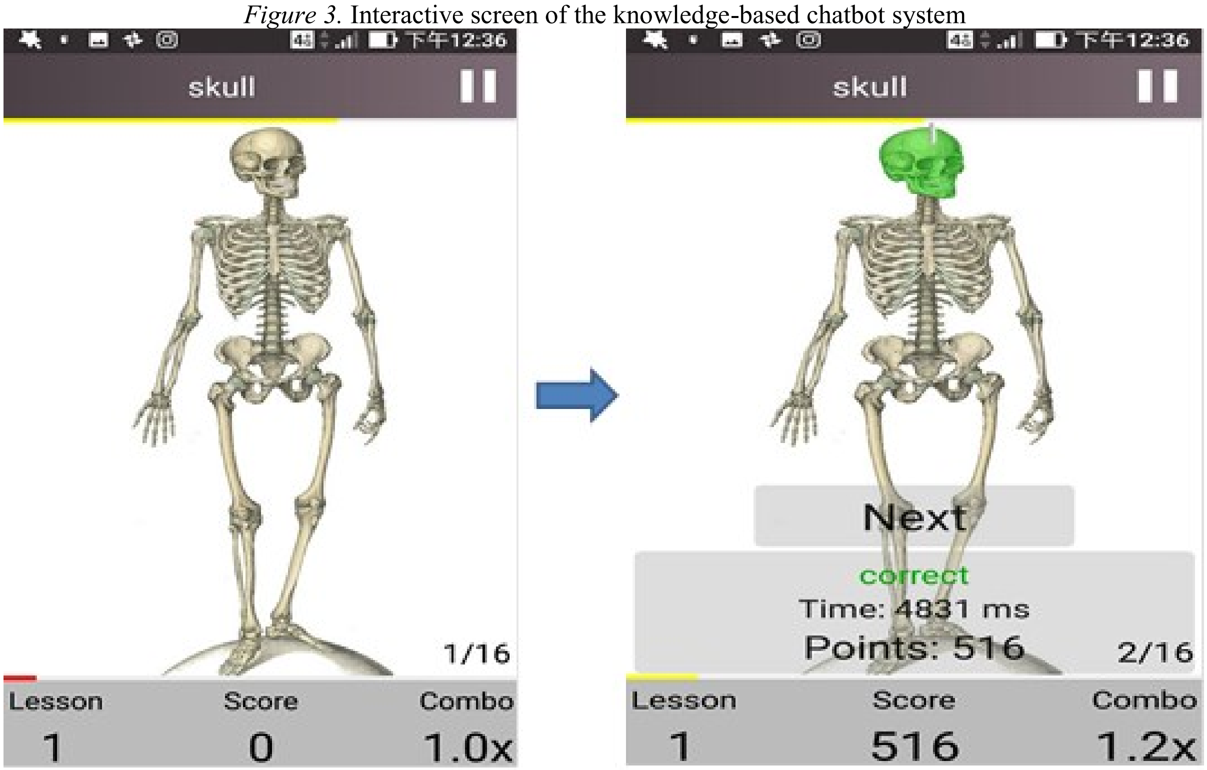
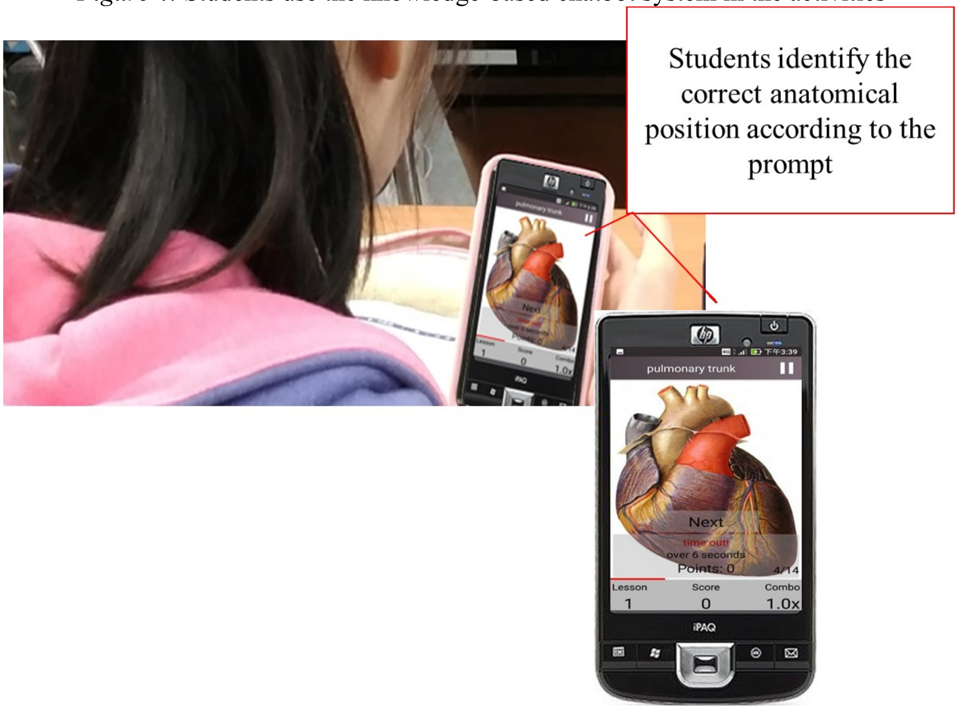


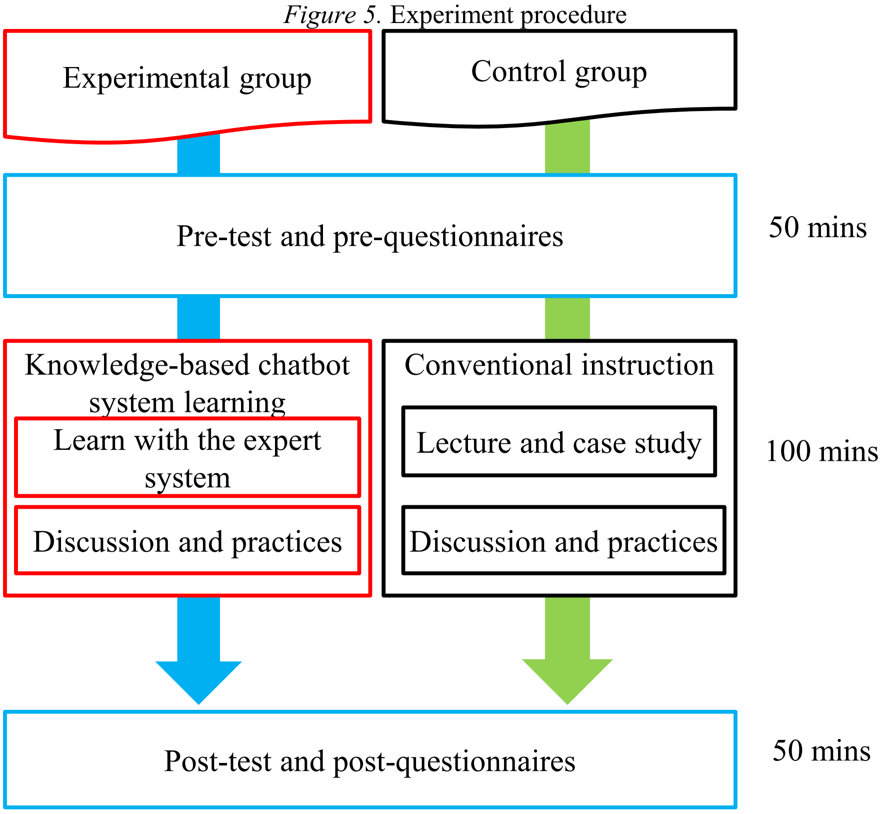
Figure 4. Students use the knowledge-based chatbot system in the activities



3.3. Experimental process

Figure 5 shows the experimental procedures illustrating the synopsis. This experiment consisted of four lessons of 50 minutes each, with a total of 200 minutes. Before the start of the experiment, the teacher introduced the physical examination course and illustrated the content of the activities. Subsequently, the students took the pre-test and completed the pre-questionnaire relating to critical thinking and learning satisfaction to measure the prior knowledge they had already learned and their feelings before the activities.

Figure 5. Experiment procedure



During the learning stage, the experimental group applied the knowledge-based chatbot system to learn the course content related to physical examination. For example, in one of the body assessment units, the cardiac assessment, the students not only need to understand the location of the four chambers of the cardiac anatomy, the location of the four chambers of the heart, the left atrium (LA), the right atrium (RA), the left ventricle (LV) and the right ventricle (RV) as well as the blood vessels, but also need to make judgments on different cases of physical examination to distinguish abnormal heart murmurs by seeking help from the knowledge-based chatbot.

On the other hand, the conventional teaching was applied to the control group, that is, the teachers illustrated the teaching content of Human Anatomy using relevant images and videos. The learning content of both groups was the same. During the stage of practice and discussion, the two groups of students could pose questions relating to physical examination and discuss them with their teachers or classmates. They were guided to use the knowledge they had learned to deal with the physical examination cases provided by the teacher; moreover, they were encouraged to discuss their case decisions or treatments, such as inspection, auscultation, percussion, and palpation, with their peers.

After the learning activity, the students completed the post-test and post-questionnaire relating to critical thinking and learning satisfaction.

3.4. Measuring instruments

This research evaluated students' nursing concepts as well as the decision-making or inference performances using 10 cases in the form of multiple-choice questions with a total score of 100. The pre-test and post-test were similar items with different cases. The questions were designed by two teachers who have taught nursing courses for more than 10 years. For example, one of the questions was related to the physical examination of the patient: "For a patient with the starting point of heart rhythmic pulsation is located at: (A) sinoatrial node (B) atrioventricular node (C) left atrium (D) left ventricle." The correct answer is (A) sinoatrial node. Another question was "During the auscultation of heart sounds, if clicks are found in the middle or late systole, which of the following conditions may be presented? (A) aortic regurgitation (B) aortic valve stenosis (C) mitral valve prolapse (D) ventricular septal defect." The correct answer is (C) mitral valve prolapse.

The critical thinking scale was proposed by Hwang and Chen (2017). It consists of five items, such as "I find myself pausing regularly to check my comprehension" and "I ask myself how well I accomplish my goals once I am finished." A 5-point Likert scoring scale was adopted and its Cronbach's α value was .83.

The learning satisfaction scale was proposed by Chu et al. (2010). It is composed of nine items, such as "The guidance provided by this system is helpful to me in observing the differences within the target learning objects." and "When using this system, I learned how to observe the target learning objects from new perspectives." A 5-point Likert scoring scale was adopted in the measure. Its Cronbach's α value was .91.

4. Experimental results

4.1. Learning achievement

This study used academic performance in the pre-test as the covariate, and academic performance in the post-test as the dependent variable. The Levene's test revealed that the homogeneity assumption was confirmed with $F(1, 30) = 0.66$ ($p > .05$). In addition, the verification did not violate the assumption of regression homogeneity ($F(1, 28) = 0.17$ ($p > .05$). ANCOVA was used for the post-hoc analysis in the scores given to the two groups. Table 1 shows the ANCOVA results of the two groups with $F(1, 29) =15.08$ ($p < .001$), indicating that the knowledge-based chatbot system (Mean = 87.90; SD =11.33) had a better effect when compared with conventional teaching (Mean = 62.32; SD = 14.95). The adjusted values of the experimental group and the control group were 86.77 and 63.45 respectively, indicating that the knowledge-based chatbot system could effectively enhance students' academic performance when compared to conventional teaching. Besides, the correlation coefficient (η^2 =0.342) was greater than 0.138, indicating that representing the knowledge-based chatbot system had a great impact on students' academic performance. In other words, the knowledge-based chatbot system could effectively enhance students' academic performance.

Table 1. ANCOVA of learning achievement by comparing the experimental and control groups

Group	N	Mean	*SD*	Adjusted mean	Std. error	*F*	η^2
Experimental group	16	87.90	11.33	86.77	3.83	15.08***	.342
Control group	16	62.32	14.95	63.45	3.83		

Note. *** $p < .001$.

4.2. Critical thinking

This study used critical thinking in the pre-test as the covariate and critical thinking in the post-test as the dependent variable. The Levene's test revealed that the homogeneity assumption was confirmed with $F(1, 30) = 0.002$ ($p > .05$). In addition, the verification did not violate the assumption of regression homogeneity ($F(1, 28) = 0.65$ ($p > .05$). ANCOVA was used for the post-hoc analysis of the scores given to the two groups. Table 2 shows the ANCOVA results of the two groups $F(1, 29) = 14.06$ ($p < .001$), indicating that the knowledge-based chatbot system group (Mean = 4.07; *SD* = 0.65) had better critical thinking when compared with the conventional teaching group (Mean = 2.83; *SD* = 0.68). The adjusted values of the experimental group and the control group were 3.99 and 2.92 respectively, indicating that the knowledge-based chatbot system could effectively enhance students' critical thinking when compared to conventional teaching. Besides, the correlation coefficient ($\eta^2 = 0.327$) was greater than 0.138, representing that the knowledge-based chatbot system had a great impact on students' critical thinking. The experimental results indicated that the knowledge-based chatbot system could effectively enhance students' critical thinking.

Table 2. Results of ANCOVA on students' critical thinking

Group	N	Mean	*SD*	Adjusted mean	Std. error	*F*	η^2
Experimental group	16	4.07	0.65	3.99	0.19	14.06***	.327
Control group	16	2.83	0.68	2.92	0.19		

Note. *** $p < .001$.

4.3. Learning satisfaction

This study used learning satisfaction in the pre-test as the covariate, and learning satisfaction in the post-test as the dependent variable. The Levene's test revealed that the homogeneity assumption was confirmed with $F(1, 30) = 0.95$ ($p > .05$). In addition, the verification did not violate the assumption of regression homogeneity with $F(1, 28) = 0.27$ ($p > .05$). ANCOVA was used for the post-hoc analysis of the scores given to the two groups. Table 3 shows the ANCOVA results of the two groups $F(1, 29) = 20.66$ ($p < .001$), indicating that the knowledge-based chatbot system group (Mean = 4.19; *SD* = 0.72) had better learning satisfaction when compared with the conventional teaching group (Mean = 2.83; *SD* = 0.68). The adjusted values of the experimental group and the control group were 4.20 and 2.83 respectively, indicating that the knowledge-based chatbot system could effectively enhance students' learning satisfaction when compared to conventional teaching. Besides, the correlation coefficient ($\eta^2 = 0.416$) was greater than 0.138, representing that the knowledge-based chatbot system had a great impact on students' learning satisfaction. The experimental results indicated that the knowledge-based chatbot system could effectively enhance students' learning satisfaction.

Table 3. Results of ANCOVA on students' learning satisfaction

Group	N	Mean	*SD*	Adjusted mean	Std. error	*F*	η^2
Experimental group	16	4.19	0.72	4.20	0.19	20.66***	.416
Control group	16	2.83	0.68	2.83	0.19		

Note. *** $p < .001$.

5. Discussion and conclusions

This research integrated a knowledge-based chatbot system into a physical examination course and used smartphones as learning devices to guide students to practice anatomy knowledge during teaching activities. The experimental results indicated that the knowledge-based chatbot system enhanced students' academic performance, critical thinking, and learning satisfaction when compared with conventional teaching.

Regarding the students' academic performance, this study found that the nursing students who used the knowledge-based chatbot system as their learning method provided evidence showing the learning effectiveness of physical examination. It is because the knowledge-based chatbot system provided an interactive learning mode that students could learn what they needed to know according to their learning progress; in other words, it provided personalized learning opportunities. The results echo those of Abubakar et al. (2019), who indicated that the enhancement of academic performance depends on the learning design. That is, if the needs of individual students can be taken care of, their learning efficacy could be improved.

In terms of critical thinking, the experimental data indicated that the implementation of the knowledge-based chatbot system can promote students' critical thinking. This finding is consistent with previous studies relating to the application of knowledge-based chatbot systems in teaching activities. For example, Goksu (2016) developed a knowledge-based chatbot system to support sex education courses for eighth-grade students, and found that it guided the students to make judgments in different scenarios. Therefore, it strengthened their critical thinking, and the students had better learning performances than those learning via conventional teaching. In the past, many scholars have mentioned that the learning mode with a combination of scenarios and guidance in nursing courses can improve students' critical thinking (Hwang & Chang, 2020; Hwang & Chang, 2021).

Referring to the students' learning satisfaction, this research has shown that participants in this study were in favor of using the knowledge-based chatbot system in learning activities. In conventional teaching mode, teachers usually give lectures using graphic pictures for illustration, but there is generally a lack of opportunities for interaction between teachers and students, not to mention the provision of instant feedback to the students. The major benefit of the knowledge-based chatbot system was possibly the provision of instant feedback that enabled the students to study efficiently according to their needs (Giraud et al., 2017). As indicated by Hwang et al. (2019), the use of an appropriate interactive learning system could improve students' learning satisfaction.

The knowledge-based chatbot system can assist users in making suitable choices and enable them to conduct systematic study with a focus on particular learning content. At the same time, the system can provide practices for different learning levels according to the students' learning progress as well as helping them identify their misconceptions during the practice. Therefore, the knowledge-based chatbot system can provide individual practice and guidance that can improve students' learning efficiency and effectiveness (Tegos & Demetriadis, 2017). In terms of this learning process, students are required to actively engage in knowledge construction while the knowledge-based chatbot system plays the role of an assistant or a learning facilitator.

The findings of the present study further echo the suggestion by Yin et al. (2020) that chatbots play many different roles in education. Many chatbot applications in education focus on the methods of analyzing and predicting students' learning behaviors. This study, however, revealed that chatbots have the ability to assist with an active learning mode. The knowledge-based chatbot system can be considered as a tutor, allowing teachers to have more time to understand students' learning problems in class, and enabling students to engage in personalized learning according to their needs. The knowledge-based chatbot system will become a "Smart Learning Partner" if students download it onto their tablets or smartphones. Thereafter, they can change their learning mode; that is, they can learn the teaching materials in a way that suits them, and repeatedly revise unfamiliar content. It makes them feel as if they have a learning partner with common learning goals, and it can enhance their cognitive development.

Despite this study having obtained the expected results, there are some limitations due to the research design and the teaching site. For instance, the objective of this experiment only focused on student midwives. It is recommended that future studies focus on students from different backgrounds and disciplines. Moreover, this study did not record students' learning processes, so it is not possible to understand the difficulties they encountered and their learning status during the process. The relatively small sample is another limitation of the present study. Based on the findings and the limitations of the present study, we recommend some suggestions for research relating to chatbots in education in the future as follows:

- It would be interesting to further investigate the chatbot system-based learning approach in relation to the learning performances and perceptions of students with different personal characteristics, such as knowledge levels, learning anxiety or self-efficacy, since the incorporation of new technologies might have different impacts on students with different personal characteristics or learning status.
- In addition to nursing students, school teachers, patients, family members of patients and nursing staff also need to continuously learn and update their knowledge. Therefore, it is important to conduct research on the benefits of using knowledge-based chatbot systems for these potential learners.
- In the traditional instruction mode, teachers generally have difficulty knowing the learning status and problems of students who need additional support. It is expected that the students' learning process can be analyzed using chatbot applications in the future. It is suggested that researchers who intend to develop chatbot applications for educational purposes not only record students' learning behaviors and status, but

also provide the logs and analysis results to teachers, such that the teachers have the opportunity to understand students' learning status and provide personal support to them, as suggested Xie et al. (2019). More importantly, the teachers would then be able to improve the learning content or learning design accordingly. That is, researchers can consider developing a class management module in chatbot-based education applications.

- A number of previous studies have mainly focused on the development of chatbot-based education applications. This may be because a majority of such studies were conducted by the researchers with a background in computer science. It is therefore recommended that cross-disciplinary research should be conducted in the future. For example, collaboration between computer science, education, and various disciplines could be extremely productive. It is expected that in-depth investigations on chatbot-based education applications can be performed. Moreover, it is also suggested that future studies can be conducted by incorporating different learning strategies (such as gamification, peer-assessment, and problem-based learning) into the learning designs using chatbot-based applications.
- It is also important to explore the long-term effects of the chatbot-based learning approach on students' learning motivation, engagement and self-efficacy as well as their learning achievements since one of the benefits of using chatbot-based applications is the provision of a personalized learning opportunity, which is related to active learning and self-directed learning.
- Similar approaches can also be employed in other nursing training programs or other fields, such as science, social science or language courses.

The major contributions of the present study are to propose a chatbot system-based learning approach and to show the effectiveness of the approach in several dimensions. The findings reported in this study could be a reference for those researchers who intend to implement research on chatbots in education as well as school teachers who intend to improve students' learning performances via the use of chatbot technologies. Moreover, in facing the recent COVID-19 problem, the use of chatbots could be a potential approach to reducing the risk of face-to-face instructions while encouraging students to explore and think in depth in professional training programs.

Acknowledgement

This study was funded by the Ministry of Science and Technology of Taiwan. The project ID was MOST-109-2635-H-227-001, MOST 108-2511-H-224-006-MY3, and MOST 110-2511-H-038-008.

References

Abbas, M. A., Hwang, G. J., Ajayi, S., Mustafa, G., & Bilal, M. (2021). Modelling and exploiting taxonomic knowledge for developing mobile learning systems to enhance children's structural and functional categorization. *Computers and Education: Artificial Intelligence*, 100007. https://doi.org/10.1016/j.caeai.2021.100007

Abbasi, S., & Kazi, H. (2014). Measuring effectiveness of learning chatbot systems on Student's learning outcome and memory retention. *Asian Journal of Applied Science and Engineering*, *3*(2), 251-260.

Abubakar, A. M., Elrehail, H., Alatailat, M. A., & Elçi, A. (2019). Knowledge management, decision-making style and organizational performance. *Journal of Innovation & Knowledge*, *4*(2), 104-114.

Balsmeier, B., Assaf, M., Chesebro, T., Fierro, G., Johnson, K., Johnson, S., Li, G.-C., Lück, S., O'Reagan, D., Yeh, B., Zang, G., & Fleming, L. (2018). Machine learning and natural language processing on the patent corpus: Data, tools, and new measures. *Journal of Economics & Management Strategy*, *27*(3), 535-553. https://doi.org/10.1111/jems.12259

Beattie, A., Edwards, A. P., & Edwards, C. (2020). A Bot and a smile: Interpersonal impressions of chatbots and humans using emoji in computer-mediated communication. *Communication Studies, 71*(3), 409-427. https://doi.org/10.1080/10510974.2020.1725082

Beaumont, I. H. (1994). User modelling in the interactive anatomy tutoring system anatom-tutor. *User Modeling and User-Adapted Interaction*, *4*(1), 21-45.

Chang, C. Y., & Hwang, G. J. (2018). Trends in smartphone-supported medical education: A Review of journal publications from 2007 to 2016. *Knowledge Management & E-Learning: An International Journal*, *10*(4), 389-407.

Chang, C. Y., Hwang, G. J., & Gau, M. L. (2021). Promoting students' learning achievement and self-efficacy: A mobile chatbot approach for nursing training. *British Journal of Educational Technology.* https://doi.org/10.1111/bjet.13158

Chang, C. Y., Lai, C. L., & Hwang, G. J. (2018). Trends and research issues of mobile learning studies in nursing education: A review of academic publications from 1971 to 2016. *Computers & Education*, *116*, 28-48.

Chang, J. H., & Tseng, C. Y. (2019). Analyzing google trends with travel keyword rankings to predict tourists into a group. *Journal of Internet Technology*, *20*(1), 247-256.

Chen, C. Y. (2012). An Innovative knowledge management learning cycle by Lego NXT for science education. *International Journal of Innovative Computing, Information and Control*, *8*(1B), 791-798.

Chen, H. L., Vicki Widarso, G., & Sutrisno, H. (2020a). A Chatbot for learning Chinese: Learning achievement and technology acceptance. *Journal of Educational Computing Research*, *58*(6), 1161-1189. https://doi.org/10.1177/0735633120929622

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. https://doi.org/10.1016/j.caeai.2020.100002

Chu, H. C., Hwang, G. J., Tsai, C. C., & Tseng, J. C. (2010). A Two-tier test approach to developing location-aware mobile learning systems for natural science courses. *Computers & Education*, *55*(4), 1618-1627.

Chou, C., Chan, P.-S., & Wu, H.-C. (2007). Using a two-tier test to assess students' understanding and alternative conceptions of cyber copyright laws. *British Journal of Educational Technology, 38*(6), 1072-1084.

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.

Darshan Singh, A., Raghunathan, S., Robeck, E., & Sharma, B. (2018). *Cases on smart learning environments*. Hershey, PA: IGI Global.

Erickson, M., & Kim, P. (2020). Can chatbots work well with knowledge management systems? *Issues in Information Systems*, *21*(4), 53-38.

Fryer, L. K., Nakao, K., & Thompson, A. (2019). Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior*, *93*, 279-289.

Gil, M., El Sherif, R., Pluye, M., Fung, B. C., Grad, R., & Pluye, P. (2019). Towards a knowledge-based recommender system for linking electronic patient records with continuing medical education information at the point of care. *IEEE Access*, *7*, 15955-15966.

Giraud, S., Brock, A. M., Macé, M. J. M., & Jouffrais, C. (2017). Map learning with a 3D printed interactive small-scale model: Improvement of space and text memorization in visually impaired students. *Frontiers in Psychology*, *8*, 930.

Goksu, I. (2016). The Evaluation of the cognitive learning process of the renewed bloom taxonomy using a web based expert system. *Turkish Online Journal of Educational Technology, 15*(4), 135-151.

Hernandez, J. P. T. (2019). Network diffusion and technology acceptance of a nurse chatbot for chronic disease self-management support: A Theoretical perspective. *The Journal of Medical Investigation*, *66*(1.2), 24-30.

Hsu, L. (2020). To CALL or not to CALL: empirical evidence from neuroscience. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2020.1750429

Hwang, G. J., & Chang, C. Y. (2020). Facilitating decision-making performances in nursing treatments: A Contextual digital game-based flipped learning approach. *Interactive Learning Environments*, 1-16. https://doi.org/10.1080/10494820.2020.1765391

Hwang, G. J., & Chang, C. Y. (2021). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2021.1952615

Hwang, G. J., Wu, Y. J., & Chang, C. Y. (2021). A Mobile-assisted peer assessment approach for evidence-based nursing education. *CIN-Computers Informatics Nursing.* https://doi.org/10.1097/cin.0000000000000753

Hwang, G. J., & Chang, H. F. (2011). A Formative assessment-based mobile learning approach to improving the learning attitudes and achievements of students. *Computers & Education, 56*(4), 1023-1031.

Hwang, G. J., & Chen, C. H. (2017). Influences of an inquiry-based ubiquitous gaming design on students' learning achievements, motivation, behavioral patterns, and tendency towards critical thinking and problem solving. *British Journal of Educational Technology*, *48*(4), 950-971.

Hwang, G. J., Chen, M. R. A., Sung, H. Y., & Lin, M. H. (2019). Effects of integrating a concept mapping-based summarization strategy into flipped learning on students' reading performances and perceptions in Chinese courses. *British Journal of Educational Technology*, *50*(5), 2703-2719.

Hwang, G. J., Sung, H. Y., Chang, S. C., & Huang, X. C. (2020a). 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*, 100003. https://doi.org/10.1016/j.caeai.2020.100003

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020b). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, *1*, 100001. https://doi.org/10.1016/j.caeai.2020.100001

Innocent, E. (2016). Trends and challenges toward integration of traditional medicine in formal health-care system: Historical perspectives and appraisal of education curricula in Sub-Sahara Africa. *Journal of Intercultural Ethnopharmacology*, *5*(3), 312–316.

Jerome, C., Cheng, J. L. A., & Hie, T. S. (2019). Undergraduate students' and lecturers' perceptions on teaching strategies that could enhance higher order thinking skills (HOTS). *International Journal of Education*, *4*(30), 60-70.

Kapočiūtė-Dzikienė, J. (2020). A Domain-specific generative chatbot trained from little data. *Applied Sciences*, *10*(7), 2221. https://doi.org/10.3390/app10072221

Kim, E., Caraballo, P. J., Castro, M. R., Pieczkiewicz, D. S., & Simon, G. J. (2019). Towards more accessible precision medicine: Building a more transferable machine learning model to support prognostic decisions for micro-and macrovascular complications of type 2 diabetes mellitus. *Journal of Medical Systems*, *43*(7), 185. https://doi.org/10.1007/s10916-019-1321-6

Kumar, B. S. (2020). A Subject-specific chatbots for primary education end-users using machine learning techniques. *International Journal of Control and Automation*, *13*(2), 407-415.

Kurniawan, M. H., & Witjaksono, G. (2018). Human anatomy learning systems using augmented reality on mobile application. *Procedia Computer Science*, *135*, 80-88.

Lan, Y. J. (2020). Immersion, interaction and experience-oriented learning: Bringing virtual reality into FL learning. *Language Learning and Technology, 24*(1), 1–15.

Lin, M. P. C., & Chang, D. (2020). Enhancing post-secondary writers' writing skills with a chatbot. *Educational Technology and Society*, *23*(1), 78-92.

Love-Koh, J., Peel, A., Rejon-Parrilla, J. C., Ennis, K., Lovett, R., Manca, A., Chalkidou, A., Wood, H., & Taylor, M. (2018). The Future of precision medicine: Potential impacts for health technology assessment. *Pharmacoeconomics*, *36*(12), 1439-1451. https://doi.org/10.1007/s40273-018-0686-6

Moore, K. L., Dalley, A. F., & Agur, A. M. (2013). *Clinically oriented anatomy*. Philadelphia, PA: Lippincott Williams & Wilkins.

Narula, J., Chandrashekhar, Y., & Braunwald, E. (2018). Time to add a fifth pillar to bedside physical examination: Inspection, palpation, percussion, auscultation, and insonation. *JAMA Cardiology*, *3*(4), 346-350.

Pérez, J. Q., Daradoumis, T., & Puig, J. M. M. (2020). Rediscovering the use of chatbots in education: A Systematic literature review. *Computer Applications in Engineering Education*. https://doi.org/10.1002/cae.22326

Pérez-Soler, S., González-Jiménez, M., Guerra, E., & de Lara, J. (2019). Towards conversational syntax for domain-specific languages using chatbots. *Journal of Object Technology*, *18*(2), Article 5-1.

Pérez-Soler, S., Guerra, E., & de Lara, J. (2018). Collaborative modeling and group decision making using chatbots in social networks. *IEEE Software*, *35*(6), 48-54.

Rather, I. A., Kim, B. C., Bajpai, V. K., & Park, Y. H. (2017). Self-medication and antibiotic resistance: Crisis, current challenges, and prevention. *Saudi Journal of Biological Sciences*, *24*(4), 808-812.

Samarakou, M., Tsaganou, G., & Papadakis, A. (2018). An e-Learning system for extracting text comprehension and learning style characteristics. *Educational Technology & Society*, *21*(1), 126-136.

Saura, J. R., Reyes-Menendez, A., & Bennett, D. R. (2019). How to extract meaningful insights from UGC: A knowledge-based method applied to education. *Applied Sciences*, *9*(21), 4603. https://doi.org/10.3390/app9214603

Segal, A., Gal, K., Shani, G., & Shapira, B. (2019). A Difficulty ranking approach to personalization in E-learning. *International Journal of Human-Computer Studies*, *130*, 261-272.

Shawar, B. A. (2017). Integrating CALL systems with chatbots as conversational partners. *Computación y Sistemas*, *21*(4), 615-626.

Sheth, A., Yip, H. Y., & Shekarpour, S. (2019). Extending patient-chatbot experience with internet-of-things and background knowledge: Case studies with healthcare applications. *IEEE Intelligent Systems*, *34*(4), 24-30.

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. https://doi.org/10.2196/14658

Shum, H. Y., He, X. D., & Li, D. (2018). From eliza to XiaoIce: Challenges and opportunities with social chatbots. *Frontiers of Information Technology and Electronic Engineering*, *19*(1), 10-26.

Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A Review of educational chatbots for the Facebook Messenger. *Computers & Education, 151*, 103862. https://doi.org/10.1016/j.compedu.2020.103862

Song, D., Rice, M., & Oh, E. Y. (2019). Participation in online courses and interaction with a virtual agent. *International Review of Research in Open and Distributed Learning*, *20*(1), 42-62.

Stuij, S. M., Drossaert, C. H., Labrie, N. H., Hulsman, R. L., Kersten, M. J., van Dulmen, S., Smets, E. M. A., & INSTRUCT project group Hanneke de Haes Arwen Pieterse Julia van Weert Noor Christoph. (2020). Developing a digital training tool to support oncologists in the skill of information-provision: A User centred approach. *BMC Medical Education*, *20*, 1-17. https://doi.org/10.1186/s12909-020-1985-0

Tegos, S., & Demetriadis, S. (2017). Conversational agents improve peer learning through building on prior knowledge. *Educational Technology and Society*, *20*(1), 99-111.

Tegos, S., Demetriadis, S., Papadopoulos, P. M., & Weinberger, A. (2016). Conversational agents for academically productive talk: A Comparison of directed and undirected agent interventions. *International Journal of Computer-Supported Collaborative Learning*, *11*(4), 417-440.

Thirumalaraju, P., Kanakasabapathy, M. K., Bormann, C. L., Kandula, H., Pavan, S. K. S., Yarravarapu, D., & Shafiee, H. (2019). Human sperm morphology analysis using smartphone microscopy and deep learning. *Fertility and Sterility*, *112*(3), e41. https://doi.org/10.1016/j.fertnstert.2019.07.237

Trasmundi, S. B., & Linell, P. (2017). Insights and their emergence in everyday practices: The Interplay between problems and solutions in emergency medicine. *Pragmatics & Cognition*, *24*(1), 62-90.

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*, 1-13.

Valtolina, S., Barricelli, B. R., & Di Gaetano, S. (2020). Communicability of traditional interfaces VS chatbots in healthcare and smart home domains. *Behaviour and Information Technology*, *39*(1), 108-132.

Van Seters, J. R., Ossevoort, M. A., Tramper, J., & Goedhart, M. J. (2012). The Influence of student characteristics on the use of adaptive e-learning material. *Computers & Education*, *58*(3), 942-952.

Wang, J., Hwang, G. H., & Chang, C. Y. (2021). Directions of the 100 most cited chatbot-related human behavior research: A review of academic publications. *Computers and Education: Artificial Intelligence*, 100023. https://doi.org/10.1016/j.caeai.2021.100023

Ward, T., Falconer, L., Frutos-Perez, M., Williams, B., Johns, J., & Harold, S. (2016). Using virtual online simulations in second Life (R) to engage undergraduate psychology students with employability issues. *British Journal of Educational Technology*, *47*(5), 918-931.

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. https://doi.org/10.1016/j.compedu.2019.103599

Xin, Y. P., Park, J. Y., Tzur, R., & Si, L. (2020). The Impact of a conceptual model-based mathematics computer tutor on multiplicative reasoning and problem-solving of students with learning disabilities. *The Journal of Mathematical Behavior*, *58*, 100762. https://doi.org/10.1016/j.jmathb.2020.100762

Yin, J., Goh, T. T., Yang, B., & Xiaobin, Y. (2020). Conversation technology with micro-learning: The Impact of chatbot-based learning on students' learning motivation and performance. *Journal of Educational Computing Research*, 0735633120952067. https://doi.org/10.1177/0735633120952067.

Yulianto, T., Andryana, S., & Gunaryati, A. (2020). Expert system for autism prediction in children with web-based forward chaining method: Expert system for autism prediction in children with web-based forward chaining method. *Jurnal Mantik*, *3*(4), 522-530.

Zhang, Y., Chen, H., Lu, J., & Zhang, G. (2017). Detecting and predicting the topic change of Knowledge-based Systems: A Topic-based bibliometric analysis from 1991 to 2016. *Knowledge-Based Systems*, *133*, 255-268.

Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two Decades of Artificial Intelligence in Education: Contributors, Collaborations, Research Topics, Challenges, and Future Directions. *Educational Technology & Society*, *25* (1), 28-47.

Two Decades of Artificial Intelligence in Education: Contributors, Collaborations, Research Topics, Challenges, and Future Directions

Xieling Chen[1], Di Zou[2*], Haoran Xie[3], Gary Cheng[1], and Caixia Liu[1,4]

[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// [4]Institute of EduInfo Science & Engineering, Nanjing Normal University, China // xielingchen0708@gmail.com // dizoudaisy@gmail.com // hrxie2@gmail.com // chengks@eduhk.hk // cxsqz@126.com

[*]Corresponding author

(Submitted January 12, 2021; Revised March 4, 2021; Accepted March 15, 2021)

ABSTRACT: With the increasing use of Artificial Intelligence (AI) technologies in education, the number of published studies in the field has increased. However, no large-scale reviews have been conducted to comprehensively investigate the various aspects of this field. Based on 4,519 publications from 2000 to 2019, we attempt to fill this gap and identify trends and topics related to AI applications in education (AIEd) using topic-based bibliometrics. Results of the review reveal an increasing interest in using AI for educational purposes from the academic community. The main research topics include intelligent tutoring systems for special education; natural language processing for language education; educational robots for AI education; educational data mining for performance prediction; discourse analysis in computer-supported collaborative learning; neural networks for teaching evaluation; affective computing for learner emotion detection; and recommender systems for personalized learning. We also discuss the challenges and future directions of AIEd.

Keywords: Artificial intelligence in education, Structural topic modeling, Bibliometric analysis, Research topics, Research evolution

1. Introduction

Artificial intelligence (AI), as a machine-based technique with algorithmic power for making predictions, diagnoses, recommendations, and decisions, has grown in importance within the educational community for its potential to support learning in diverse contexts in recent years (Hwang et al., 2020a). The field of AI in education (AIEd) has demonstrated technological advances, theoretical innovations, and successful pedagogical impact (Roll & Wylie, 2016), with diverse applications such as intelligent tutors for content delivery, feedback provision, and progress supervision (Bayne, 2015). The affordances of AIEd are widely recognized. AI can be used to provide specialized support and raise knowledge-gap awareness, which enables instructors to teach effectively and efficiently through personalized and adaptive instruction (Guan et al., 2020). AI also provides algorithm-based decisions which enable effective real-time assessment of complex skills and knowledge (Chen et al., 2021). Additionally, AI-empowered educational systems can be used to analyze classroom dynamics and student engagement, which in turn helps to identify at-risk students in real-time mode, thus enabling timely intervention (Tsai et al., 2020).

Researchers and practitioners have been promoting AI and exploiting its pedagogical potential; consequently, scientific output on AIEd has increased significantly (Hinojo-Lucena et al., 2019). Scientific literature is valuable for thoroughly understanding the history and status of a field and can be analyzed through research motivation identification, scientific collaboration evaluation, and research theme detection (Chen et al., 2020a). Given the rapid growth of AIEd research, a synthesis of the extant literature for a summarized overview appears timely.

Several reviews that applied narrative synthesis or the systematic review of small samples have been conducted. Chassignol et al. (2018) reviewed AIEd literature from four perspectives, i.e., personalized instructional materials, innovative instructional strategies, technology-assisted assessment and communications between learners and instructors, based on 47 publications in the *International Journal of Artificial Intelligence in Education* (IJAIED) in 1994, 2004, and 2014. Roll and Wylie (2016), who explored AIEd's strengths and opportunities, found there was an evolutionary process regarding in-class learning practices and interactions with instructors supported by diversified AI technologies and a revolutionary process regarding AI technologies' adoption in students' daily life and community activities. Zawacki-Richter et al. (2019) systematically reviewed 146 publications about AI in higher education, identifying AI's applications for profiling and prediction, assessment and evaluation, adaptivity and personalization, and intelligent tutoring systems (ITSs) to support academic, institutional, and administrative services.

There are reviews on AIEd based on quantitative methodologies. Goksel and Bozkurt (2019) adopted social network analysis in reviewing AIEd publications from 1970 to 2018. They identified three themes, i.e., adaptivity/personalization and learning styles; expert systems and ITSs; and AI as an integrated component during instruction. Hinojo-Lucena et al. (2019) bibliometrically analyzed 132 AIEd publications from 2007 to 2017; their review showed there was a global interest in AIEd, and the period represented an incipient stage for publications in the area. Chen et al. (2020b) reviewed 45 AIEd-related publications in terms of annual distribution, major journals, institutions, countries/regions, research issues, and the theories and technologies involved in order to highlight gaps in AIEd applications and theory. Guan et al. (2020) analyzed 400 articles on AI and deep learning (DL) in education through manual coding and keyword analysis. Their review indicated increasing interest in implementing and designing online education from 2000 to 2009 and the prevalence of personalized learning supported by learner profiling and learning analytics (LA) from 2010 to 2019. Tang et al. (2021) systematically reviewed publications about the application of AI in e-learning, focusing on leading journals, countries, disciplines, and applications, with a co-citation network analysis examining relations among core-cited references to predict future research directions. Their review revealed that AI-based personalized learning scenarios and student characteristic prediction using Bayesian networks were prevalent. These reviews, however, have mostly adopted qualitative methods, with limited studies analyzed and specific results discussed, failing to present a thorough understanding of the general field, particularly about research topics and topic evolution. Such traditional analysis of the full contents of a publication through manual coding and synthesis, however, is time-consuming and laborious, and as the published literature rapidly increases, is becoming outmoded.

Given the prevalence of AIEd and the lack of a quantitative analysis of its copious literature, a review providing a comprehensive understanding of AIEd using rigorous machine learning (ML) appears timely. Owing to ML's rapid development, diverse approaches capable of analyzing large volumes of data are now available, among which topic models are effective and efficient for inferring latent topics from large amounts of literature (Chen et al., 2020a). The inferred information reveals a better understanding of historical and extant research progress, development of technologies applied, and drivers of fresh ideas, all of which can help researchers and educators decide upon research topics and project planning.

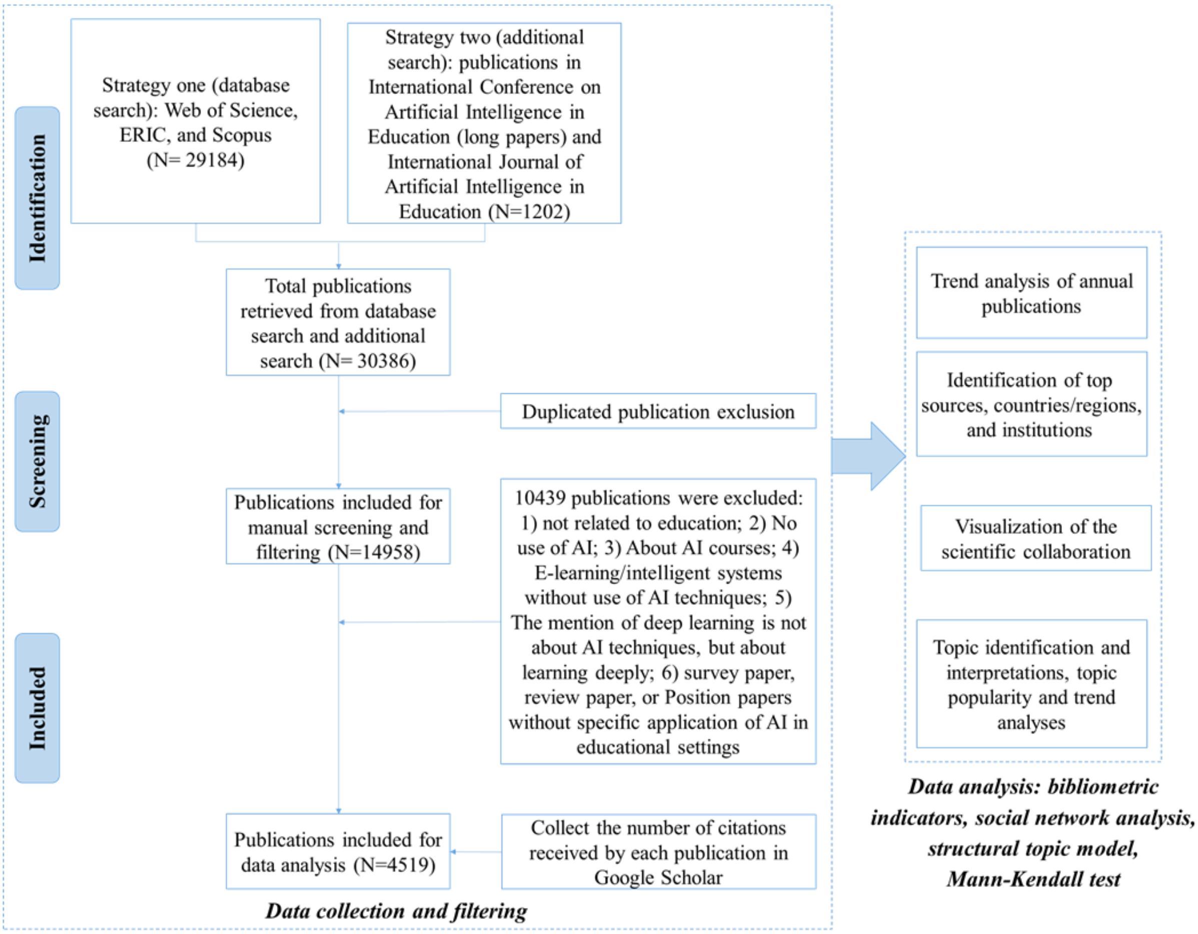
Accordingly, we applied topic-based bibliometrics to quantitatively examine 4,519 AIEd literature from 2000 to 2019 to uncover topic trends and predict the future of AIEd, focusing on the following: changes in topic popularity; major publication sources, countries/regions and institutions; and scientific collaborations. Our review was guided by five research questions:

RQ1: What were the number of AIEd articles published from 2000 to 2019?
RQ2: What were the top publication sources, countries/regions, and institutions?
RQ3: What was the nature of collaboration among countries and institutions?
RQ4: What were the most investigated research topics?
RQ5: How did the intensity of research interest in these topics change?

2. Dataset and methods

Figure 1 depicts the steps of data collection and analysis. Detailed descriptions follow:

Figure 1. Data collection and analyses



2.1. Data retrieval and preprocessing

AIEd-related publications from 2000 to 2019 were collected on May 30th, 2020 using two strategies. First, Web of Science (WoS), Scopus, and Education Resources Information Center (ERIC) databases were searched. Two lists of search terms were considered, including AI-related terms ("artificial intelligence," "machine intelligence," "intelligent support," "intelligent virtual reality," "chat bot*," "machine learning," "automated tutor*," "personal tutor*," "intelligent agent*," "expert system*," "neural network*," "natural language processing," "chatbot*," "intelligent system*," and "intelligent tutor*") and education-related terms ("education," "college*," "undergrad*," "graduate," "postgrad*," "K-12," "kindergarten*," "corporate training*," "professional training*," "primary school*," "middle school*," "high school*," "elementary school*," "teaching" and "learning"). Specifically, in WoS, "TS" was searched with AI-related terms to include research articles and conference papers written in English, and these were categorized in *Education Educational Research*. In Scopus, "TITLE-ABS-KEY" was searched with AI- and education-related terms to include articles published in journals and conference proceedings, written in English, categorized in Social Sciences and further restricted to publication sources with "education*," "teaching," "learning," or "instruction*" in their names. In ERIC, titles and abstracts were searched using individual AI-related terms, with the results being aggregated and duplicated. The first strategy identified 29,184 publications.

Second, considering the close relevance of the International Conference on Artificial Intelligence in Education (ICAIED) and IJAIED to our research target, we conducted an additional search in these two and obtained 1,202 publications.

The 30,386 publications were checked for duplication via title comparison by calculating string similarity using the Python package called *strsim*. After calculation, titles of publications with a similarity degree equal to "1" were duplicated, with the rest being sorted in descending order of similarity values for manual checking. Specifically, for the 29,184 publications retrieved using strategy one, title comparisons of Web of Science and ERIC, Web of Science and Scopus, and ERIC and Scopus were conducted to eliminate duplications. Thereafter, titles of the remaining publications were compared against the 1,202 publications retrieved using strategy two to delete duplications, resulting in 14,958 publications for further data screening.

Two domain experts adopted the criteria in Figure 1 to determine publication relevance. They each assessed the same 300 randomly selected articles independently, leading to inter-rater reliability of 91%, with inconsistencies being discussed to resolve differences. Thereafter, they screened the remaining data separately, resulting in 4,519 eligible publications whose numbers of citations were collected in Google Scholar (see https://scholar.google.com/).

Preprocessing included manual supplementation of publication features, including the author's address information by referring to original full-texts and the identification of authors' institutions and their corresponding countries/regions. To analyze research topics, terms were extracted from titles, abstracts, and keywords with a weighting strategy (Chen et al., 2018). Additionally, term frequency-inverse document frequencies with a threshold of 0.05 was conducted for term selection.

2.2. Data analysis

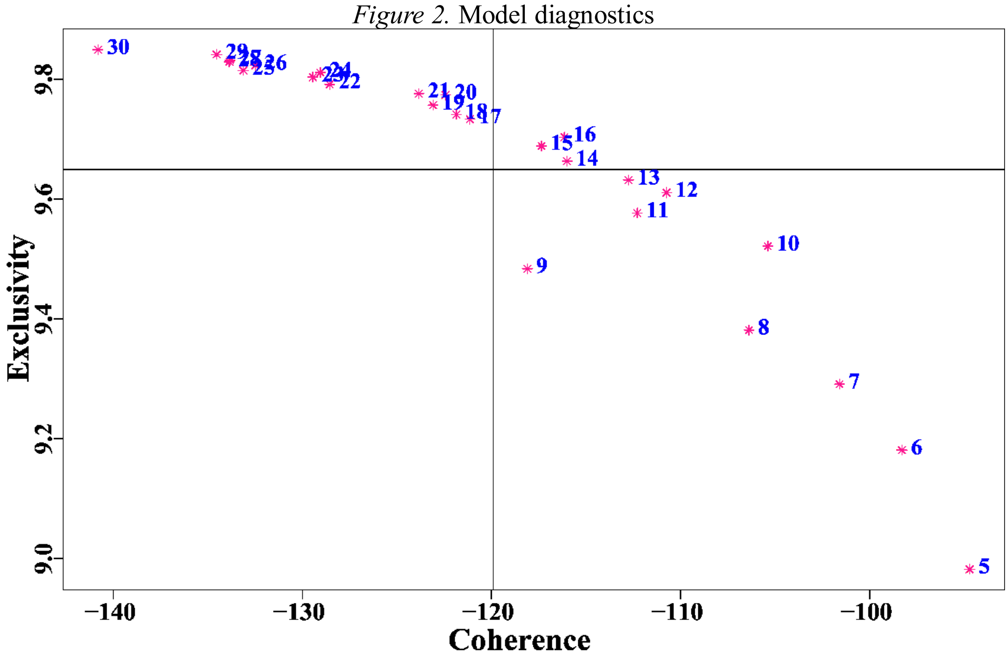
Four methodologies (i.e., bibliometric indicators, social network analysis, structural topic modeling (STM), and Mann-Kendall (MK) trend test) were used.

First, the publication count measured annual productivity. A polynomial regression analysis was further conducted to determine the developmental trend of AIEd research. Publication sources, countries/regions, and institutions were analyzed using publication count and the Hirsch index (H-index) to measure productivity and impact.

Second, social network analysis via Gephi (see https://gephi.org/) visualized relationships between institutions or countries/regions by treating institutions or countries/regions as nodes with the node size indicating their productivity and the link width indicating collaboration intensity.

Additionally, research topics in the 4,519 publications were identified using STM (Roberts et al., 2014; Roberts et al., 2019). We ran 26 models with the number of topics ranging from five to 30. Three models with 14, 15, and 16 topics each achieved higher values of semantic coherence and exclusivity measures (see Figure 2). For them, two domain experts conducted comparisons by examining representative terms and studies. The model with 16 topics (i.e., 16-topic model) was identified as it produced "the greatest semantic consistency within topics and exclusivity between topics (Chen et al., 2020a, p. 4)." To examine how the intensity of research interest in each topic changed over time, we employed the MK test (Mann, 1945).

Figure 2. Model diagnostics



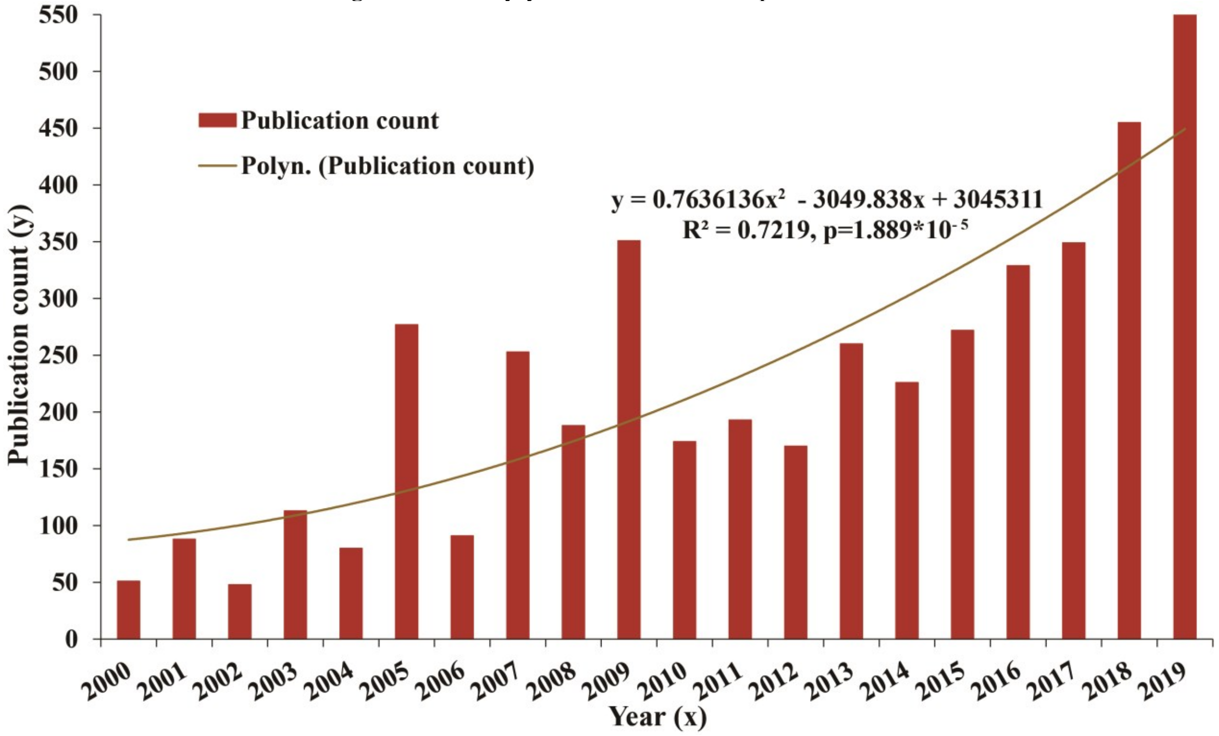
Note. Each node represents a topic model with blue labels indicating the number of topics.

3. Results

3.1. Annual numbers of AIEd publications

Figure 3 shows the number of AIEd articles published from 2000 to 2019, indicating an overall increasing tendency, particularly since 2012. The increasing interest in AIEd research is mainly due to the increased positive findings of AI's effects on learning performance and outcomes.

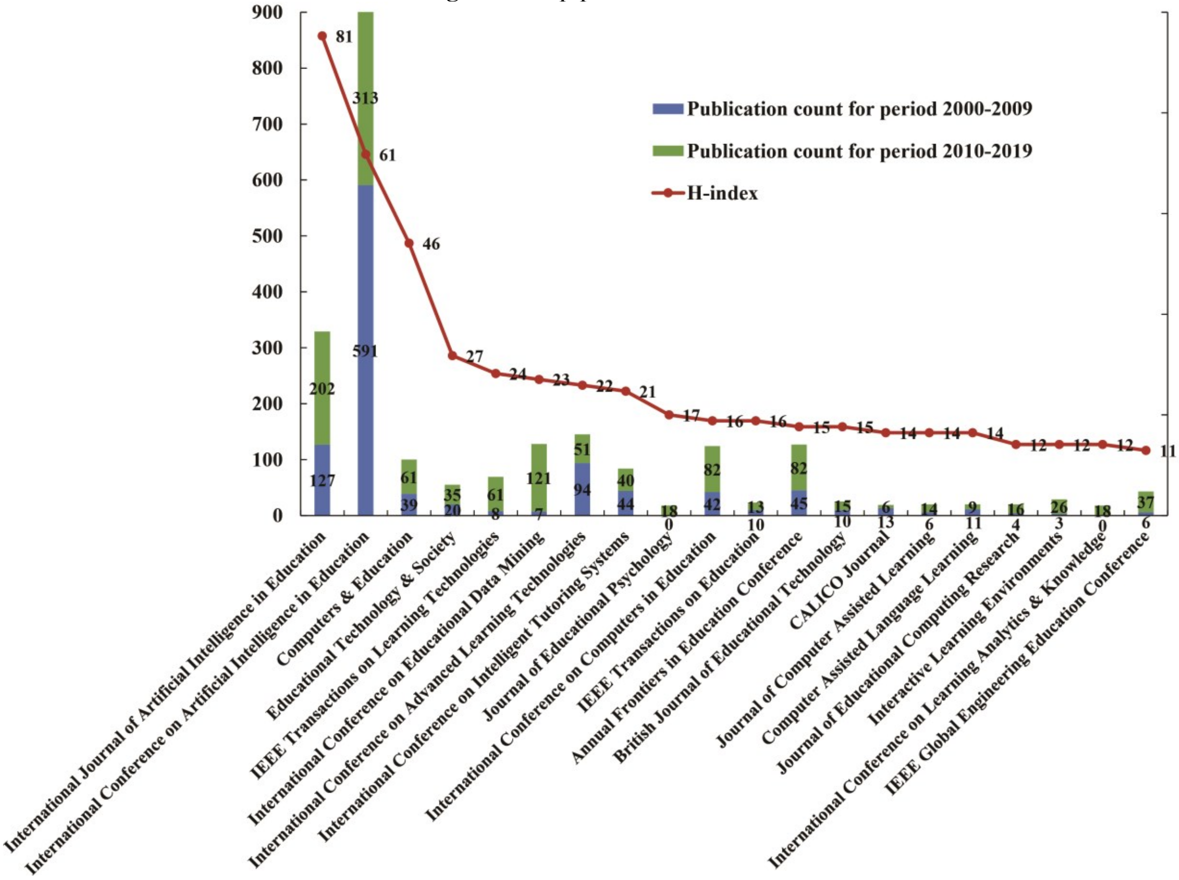
Figure 3. Year-by-year number of AIEd publications



3.2. Top publication sources

In total, 650 sources were identified, with the top 20 ranked by H-index (Figure 4) contributing to over 50% of the total. Eight were conferences, with IJAIED at the top with an H-index of 81 and 329 publications, followed by ICAIED, *Computers & Education*, and *Educational Technology & Society*. Comparing the publication counts of the first decade with the second, most sources became increasingly interested in AIEd in the latter.

Figure 4. Top publication sources



3.3. Top countries/regions and institutions

In total, 92 countries/regions were identified, with the top 20 ranked by H-index (Figure 5). The USA was at the top with 1,700 publications, 54,344 citations, and an H-index of 102. Based on the H-index, other important countries/regions included Canada, the UK, and Taiwan. We identified 2,296 institutions (top 20 in Figure 6), with Carnegie Mellon University, the University of Pittsburgh, and the University of Memphis holding the top three positions. Measured by publication count, the top three were Carnegie Mellon University, Arizona State University, and the University of Pittsburgh. Most countries/regions and institutions became increasingly interested in AIEd over the period.

Figure 5. Top countries/regions

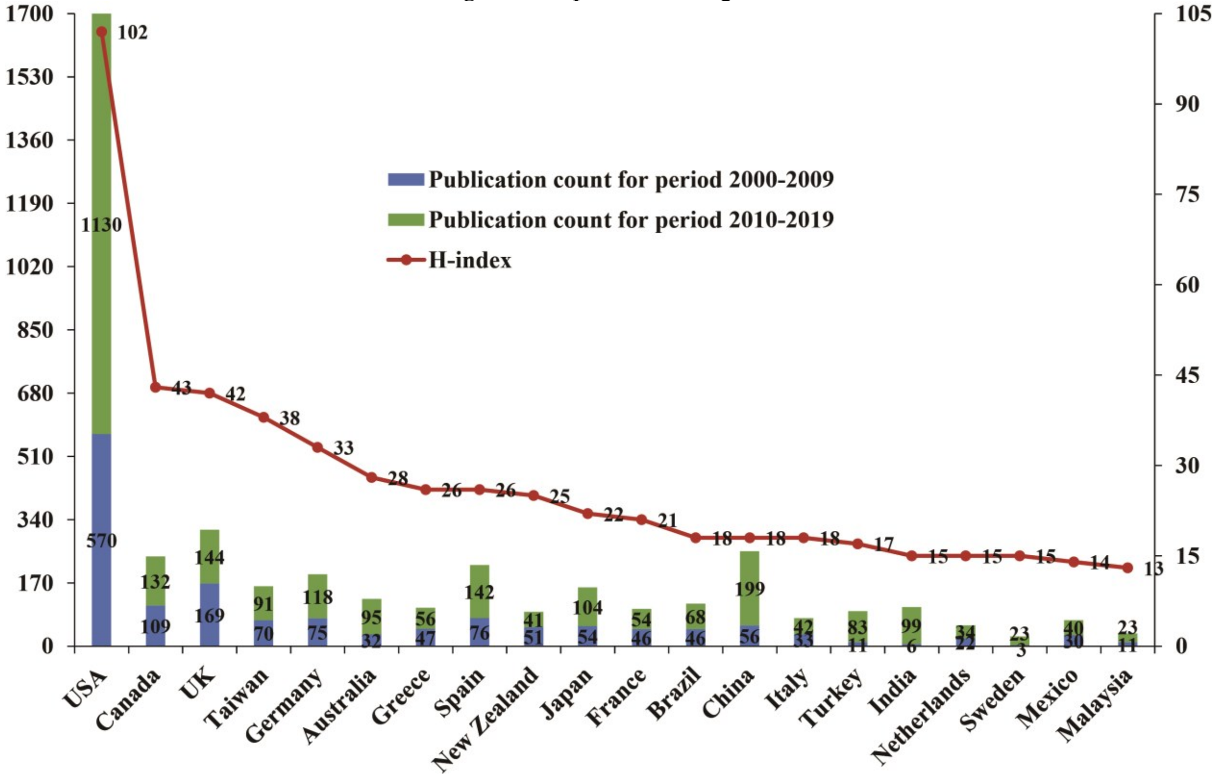
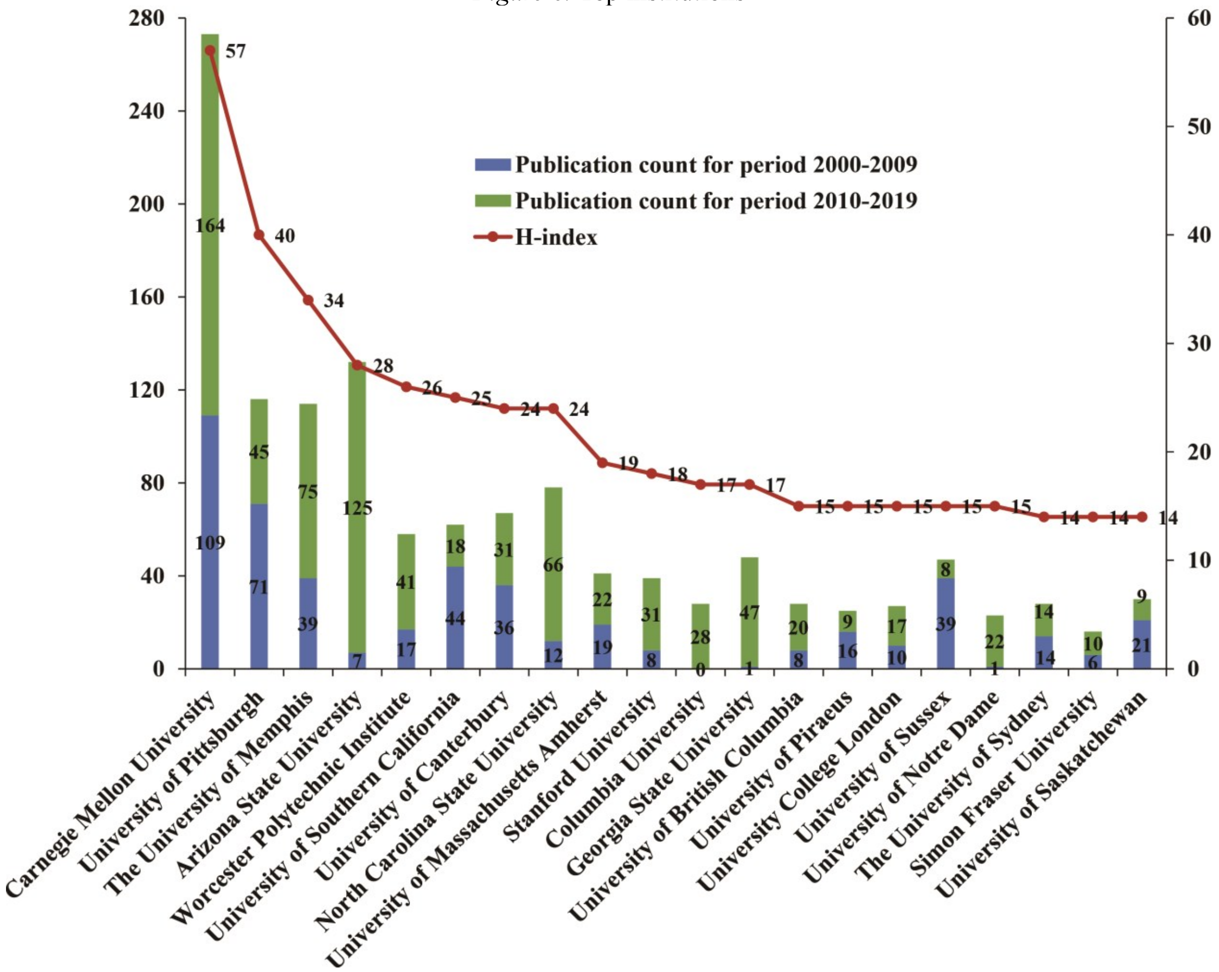


Figure 6. Top institutions



3.4. Scientific collaborations

Collaborations among the top countries/regions are visualized in Figure 7. The USA, the UK, Canada, Spain, and Australia were the most collaborative, with the USA and Germany being the closest partners. From an institutional perspective (Figure 8), Carnegie Mellon University, Arizona State University, and the University of Southern California were the most collaborative, with Georgia State University and Arizona State University being the closest partners.

Figure 7. Collaborations among top countries/regions

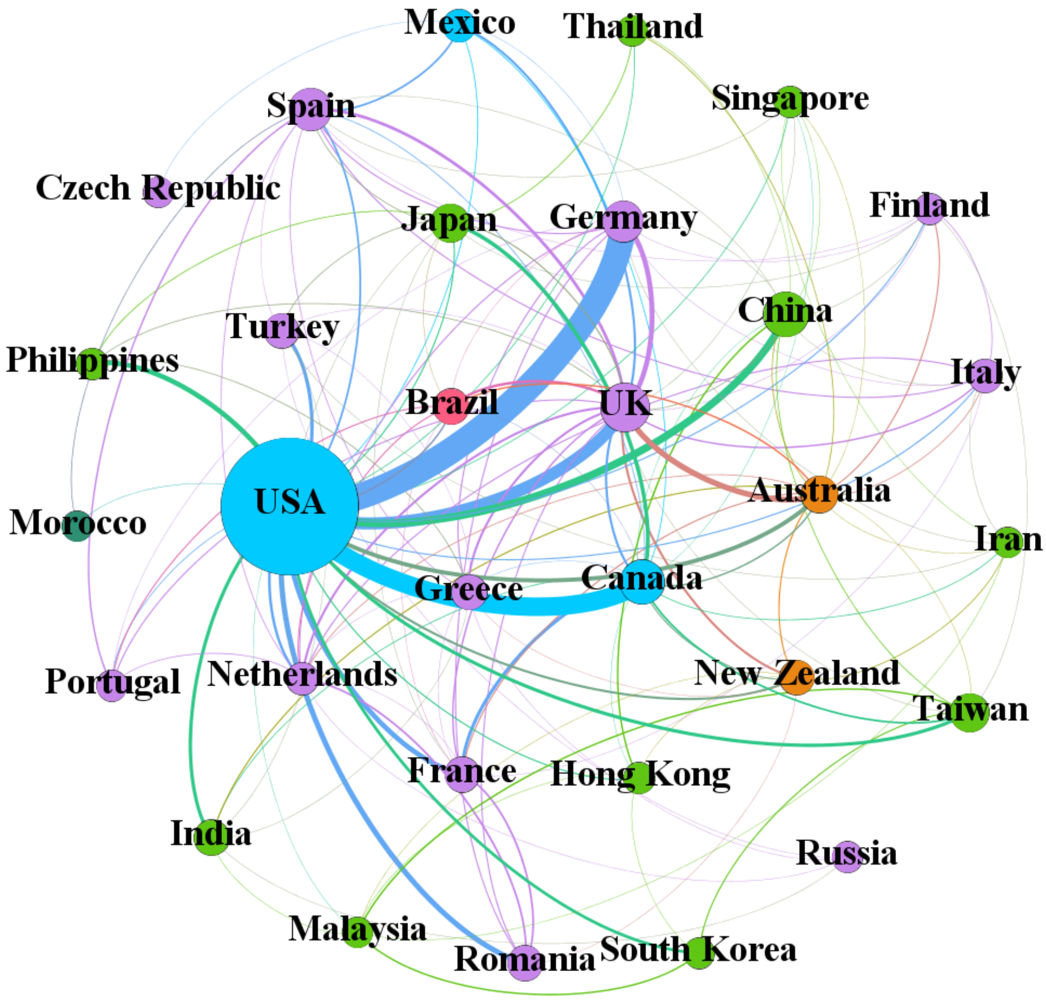
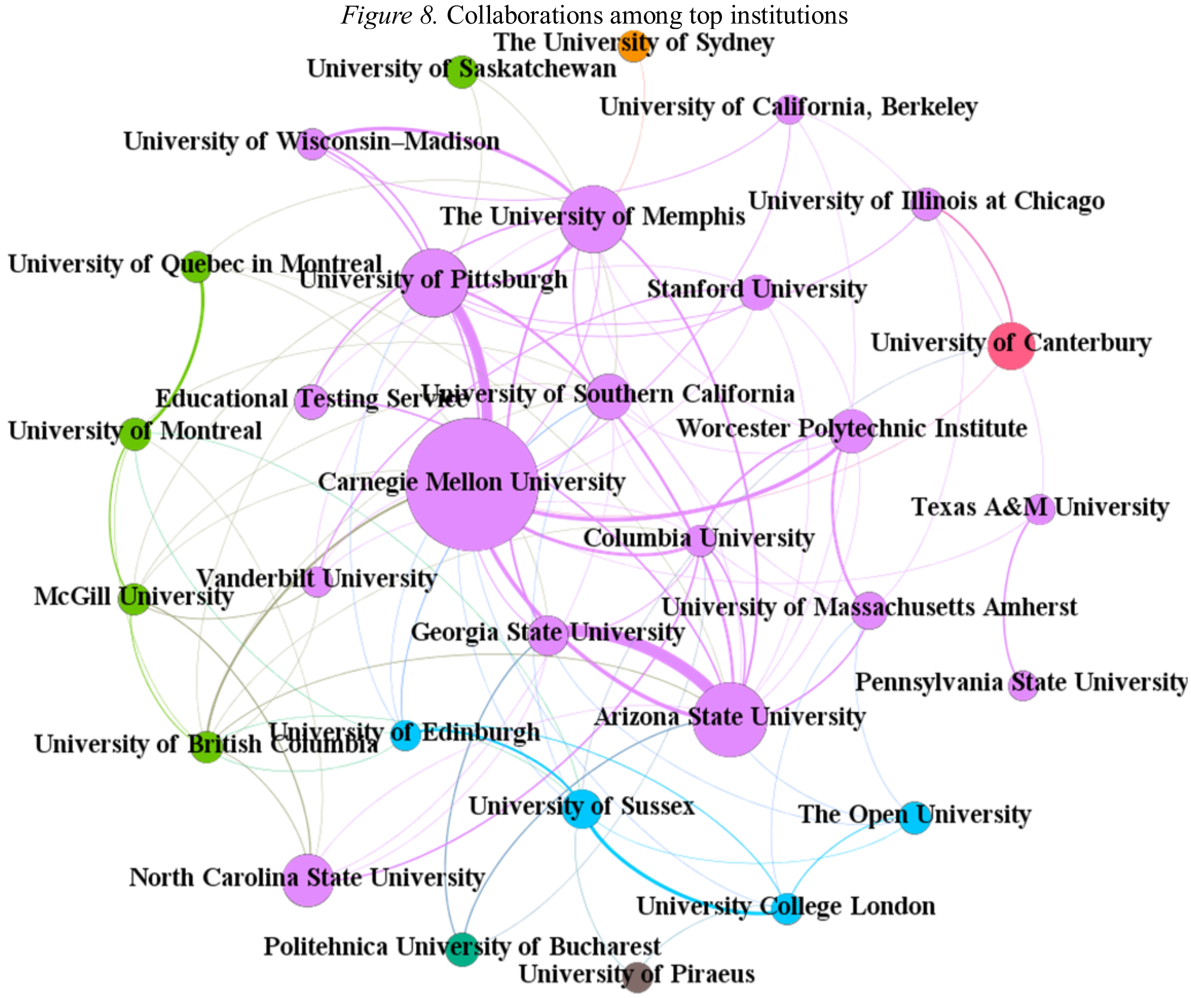


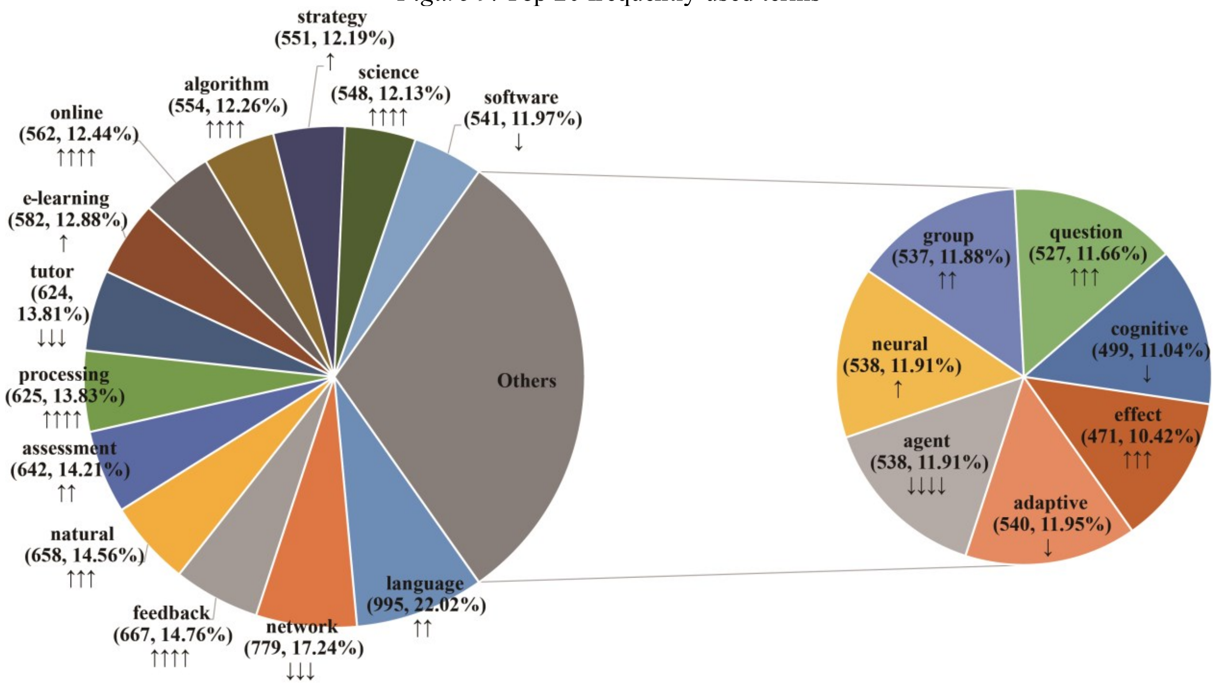
Figure 8. Collaborations among top institutions



3.5. Most frequently-used terms

Figure 9 shows the top 20 most frequently-used terms, with "language" at the top appearing in 995 publications. Other important terms included "network," "feedback," "natural" and "assessment." The trend test indicated that terms like "language," "feedback," "natural," "assessment," "processing," "online," "science," "group" and "question" experienced significant increases over the period.

Figure 9. Top 20 frequently-used terms

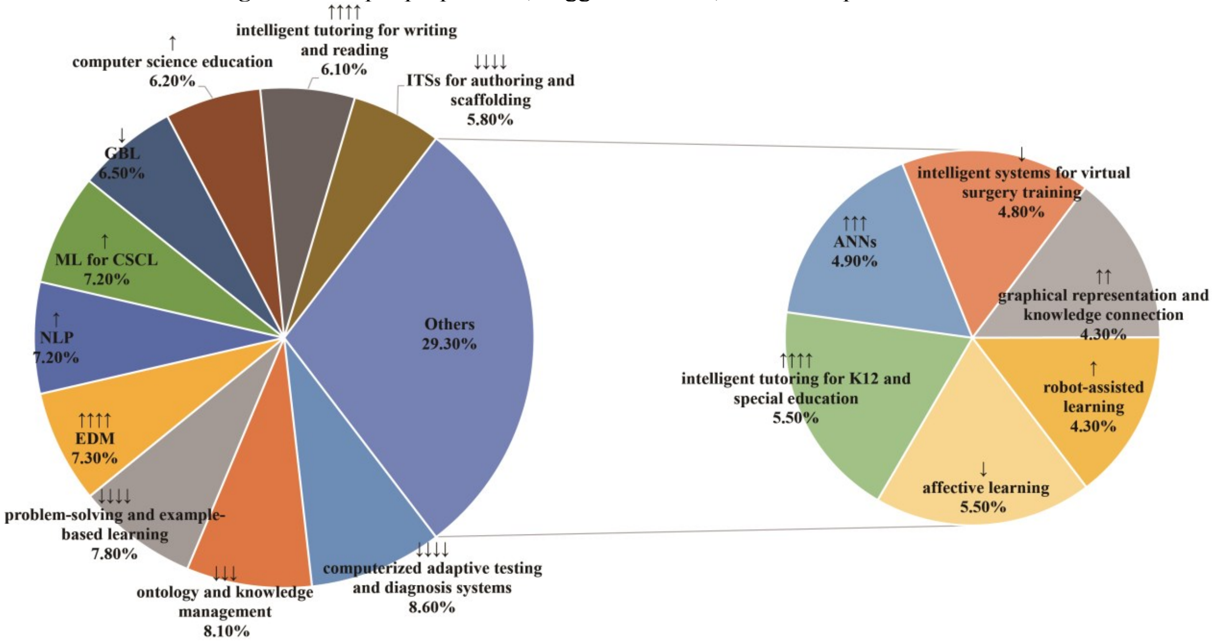


Note. inside the parentheses are term occurrence and proportion; ↑(↓): increasing (decreasing) trend but not significant ($p > .05$); ↑↑(↓↓), ↑↑↑(↓↓↓), ↑↑↑↑(↓↓↓↓): significantly increasing (decreasing) trend ($p < .05$, $p < .01$, and $p < .001$, respectively)

3.6. Research topics and topic trend

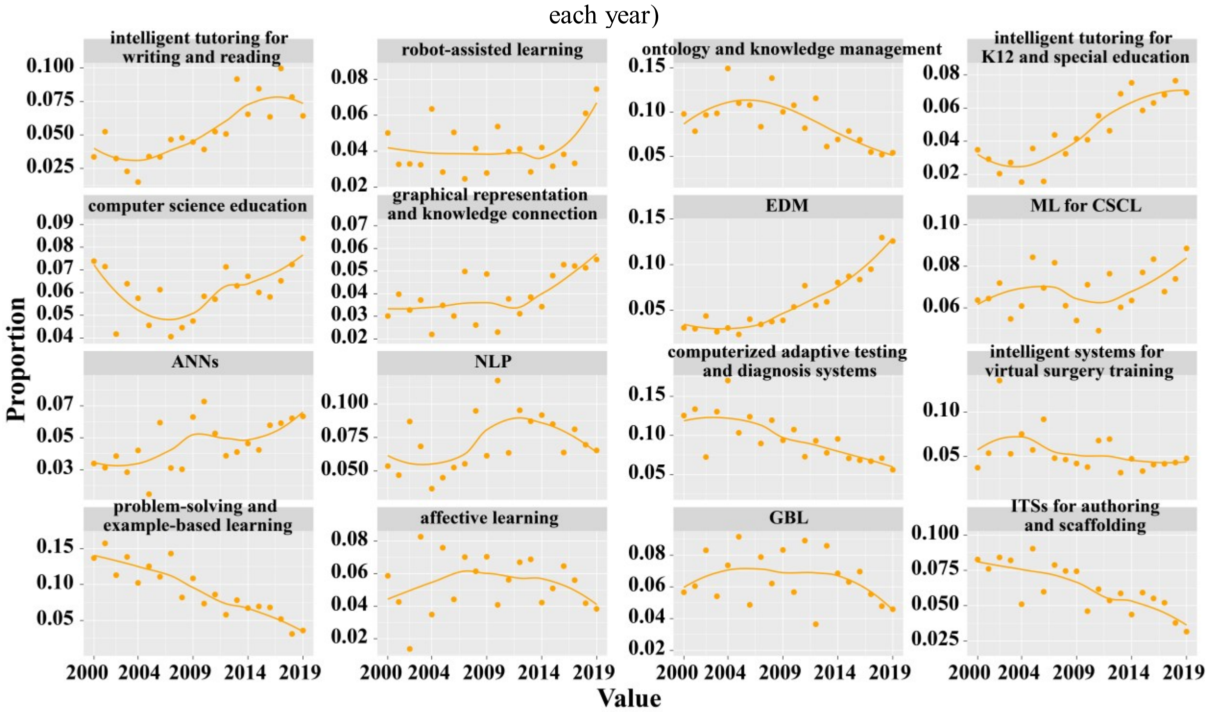
Figure 10 shows the results of the 16-topic model together with suggested labels, topic proportion, and trend test results. Five topics (i.e., *educational data mining* (*EDM*), *intelligent tutoring for writing and reading*, *intelligent tutoring for K12 and special education*, *artificial neural networks* (*ANNs*), and *graphical representation and knowledge connection*) enjoyed a significant increasing trend, whereas four topics (i.e., *computerized adaptive testing and diagnosis systems*, *ontology and knowledge management*, *problem-solving and example-based learning*, and *ITSs for authoring and scaffolding*) experienced a significant decreasing trend over the two decades. Figure 11 illustrates the annual topic proportion, indicating how popular each topic was in each year. Specifically, in the early years, AIEd scholars focused mainly on ontology use and knowledge management in ITSs to facilitate problem-solving and example-based learning for scaffolding purposes where computerized testing and diagnosis of learner knowledge and learning processes were frequently concerned. In the later years, articles on learner affect and emotion in diverse scenarios became more frequent, especially in GBL, where learners commonly experienced diverse emotions that directly impacted learning performance. Also, ITSs gradually extended their applications to facilitate the learning of diverse subjects, particularly NLP-assisted language education and in K12 and special education. Furthermore, increasingly diverse technologies were used for various educational goals, e.g., robot-assisted computer science education, ML-assisted CSCL, and ANN-assisted learning prediction and teaching evaluation. Additionally, EDM and LA were increasingly applied to visualize the learning process and knowledge acquisition for easy understanding. These foci in the later years point towards the future and challenges faced by AIEd scholars.

Figure 10. Topic proportions, suggested labels, and developmental trends



Note. % indicates topic proportion; ↑(↓), ↑↑(↓↓), ↑↑↑(↓↓↓), ↑↑↑↑(↓↓↓↓) are similar to *Figure 9*

Figure 11. Annual topic proportion (x-axis indicates publication year and the y-axis indicates topic popularity in each year)



4. Discussion

Focusing on the research questions, this section discusses the findings. For RQ1, consistent with previous reviews (Chen et al., 2020b; Hinojo-Lucena et al., 2019; Roll & Wylie, 2016; Tang et al., 2021; Zawacki-Richter et al., 2019), the overall growth of AIEd literature indicates a positive future with an expanding community and scientific output. Responding to RQ2, AIEd research is especially welcomed by interdisciplinary journals such as *Computers & Education* and *Educational Technology & Society* with their dual foci on education and technology; these journals are also highly ranked in publishing AI in e-learning studies (Tang et al., 2021). The results support Zawacki-Richter et al. (2019) and Tang et al. (2021), who highlighted AIEd's close relationship with computer science and software engineering. Consistent with Hinojo-Lucena et al. (2019) who identified the

USA as the dominant actor, our study further revealed that scholars in a variety of countries/regions (e.g., Canada, the UK, and Taiwan) and institutions were increasingly interested in AIEd. The higher AIEd research productivity in these countries/regions can be partially attributed to their governments' efforts to promote technology-enhanced learning through educational policy and funding (Chen et al., 2020c). In Tang et al., Taiwan was the top country, whereas, in our study, it was ranked 7th, which may reflect our wider focus on AI's application in education as a whole, rather than in e-learning alone. Carnegie Mellon University was the top in research productivity and impact. Responding to RQ3, the network visualization (Figures 7 and 8) revealed that the countries/regions and institutions that had intense scientific collaborations showed higher productivity and wider impact. We thus call for enhanced international collaborations to better embrace challenges as AIEd advances. Additionally, AIEd's interdisciplinarity was uncovered by the topic modeling, demonstrating effective and important AI technologies that originated from computer science.

The STM results respond to RQ4, revealing frequently occurring issues throughout the review period. These include computerized adaptive testing, diagnosis, and instruction systems integrated with varied AI technologies, especially NLP, ontology, ML, and ITSs. All of these facilitate diverse educational goals such as subject knowledge (e.g., language skills and programing) and ability (e.g., problem-solving) acquisition and innovative pedagogical strategy implementation (e.g., GBL and example-based learning). Consistent with several reviews (Chassignol et al., 2018; Guan et al., 2020; Tang et al., 2021; Zawacki-Richter et al., 2019) that identified the important roles of ITSs and AI in assessment, feedback, and learner performance prediction, we highlighted ITSs' popularity in various domain-specific types of education (e.g., K-12 education and language education), AI for computerized adaptive testing and diagnosis, and learner performance tracking and prediction using EDM. Similar to Roll and Wylie (2016) who highlighted AI's role in supporting collaborations in interactive learning, we identified CSCL assisted by ML, a technique also identified in Chen et al. (2020b). Consistent with Chassignol et al. (2018) who identified educational robots, Guan et al. who identified educational games and teaching evaluation, and Tang et al. (2021) who noted Bayesian networks and neural networks for learner learning characteristic prediction, we highlighted robot-assisted learning, GBL, and neural network-assisted teaching evaluation. Similar to Chen et al. (2020b) who identified NLP, we further highlighted its importance in language education. Just as several reviews (Chassignol et al., 2018; Guan et al., 2020; Tang et al., 2021; Zawacki-Richter et al., 2019) have identified the growing interest in AI-assisted personalization, we also noted the interest in personalization in adaptive testing and diagnosis. Roll and Wylie and Tang et al. (2021) highlighted an increasing interest in domain-level learning such as language and medical education and STEM (science, technology, engineering, and mathematics) education assisted by AI; we also revealed AI's use in various subjects and domains (e.g., computer science education, language education, K12 and special education, and surgery training). We additionally identified new topics such as problem-solving, example-based learning, authoring and scaffolding, and affective learning.

The findings of the topic analyses, and especially the trend analysis, answers RQ5, revealing there was a decreased interest in ITSs for authoring and scaffolding, whereas ITSs were increasingly used for NLP-assisted language education and K12 and special education. As for AI technologies, ontology use declined, whereas advanced techniques such as ML, ANNs, EDM gained popularity in scenarios such as CSCL; however, these were less popular in problem-solving and example-based learning. Compared to computerized testing and diagnosis, how AI facilitates subject knowledge acquisition became prevalent over the review period (e.g., robot-assisted programming education). These findings bring insight into important issues and the potential future directions of AIEd. We established eight themes by examining and interpreting the topics receiving increasing interest. For example, when considering the most representative studies of the topic *EDM* centering on EDM-assisted learning prediction, we formed a theme called "EDM for performance prediction." "NLP for language education" was established by examining three topics (i.e., *NLP*, *intelligent tutoring for writing and reading*, and *graphical representation and knowledge connection*) whose representative studies focused mainly on NLP use in language education. Other themes were formed similarly, except for "Affective computing for learner emotion detection" and "Recommender systems for personalized learning." The former was selected based on topic *affective learning*, which, although not found to increase significantly in popularity, was widely reported to facilitate instruction, particularly regarding personalization. The latter theme was included because of its prevalence in the data corpus. Although it was not identified as a separate topic due to topic overlaps in topic models, personalized learning was increasingly prevalent (e.g., Chassignol et al., 2018, Zawacki-Richter et al., 2019; Guan et al., 2020), particularly in the form of personalized material recommendation. Hereafter, tightly aligning to the eight themes, we discuss AIEd's challenges and the future effort needed to advance the field.

4.1. ITSs for special education

As an adaptive instructional system incorporating AI into educational methods, ITSs have been widely applied in various domains (e.g., STEM education, computer science education, and language education) with benefits and positive effects well documented. Instead of repeating what has been found in previous reviews, we would like to highlight an emerging need for the application of ITSs in special education, particularly among autistic students. ITSs' effectiveness for teaching autistic students owes much to their ability to provide immediate and personalized instruction and feedback, which is as effective as one-to-one tutoring. This overcomes the difficulties in anticipating and recognizing autistic students' negative behaviors (Mondragon et al., 2015). An integrative specialized learning application (ISLA) (Mondragon et al., 2016) can help autistic students manage emotions using learning trace analysis and learning performance evaluation. In ISLA, a virtual agent named Jessie adjusts an autistic learner's emotional state in real-time and provides personalized encouragement and support to assist problem-solving during learning. This feedback relieves the autistic learners' anxiety and frustration while keeping them engaged.

ITSs also benefit autistic learners in performing real-time learning tasks by monitoring and intervening when necessary. An intelligent LEGO tutoring system (Sun & Winoto, 2019) assists both instructors and autistic learners in brick playing. In the instructors' module, instructors design a new model of LEGO bricks; thereafter, visual and auditory step-by-step instructions for model completion are automatically generated. In the learners' module, the designed model is loaded with displayed instructions. Such systems benefit autistic learners by prompting instructors with necessary interventions and instructions while tracking learners' brick-building process in real-time with feedback and suggested corrections automatedly provided when a mistake is made.

4.2. NLP for language education

NLP is instrumental for computer-assisted language learning (CALL). First, many new CALL applications integrate various automatic speech recognition technologies to create realistic and engaging learning experiences by enabling computers to understand learners' speech and react accordingly or provide feedback on speech quality (Zhang & Zou, 2020). One recent call to researchers is to develop speech-to-text algorithms enabling seamless integration of speech recognition systems to enhance learners' real-time understanding of their adopted reading strategies for oral self-explanations on a given text (Panaite et al., 2018).

Second, word sense disambiguation facilitates effective vocabulary learning by resolving lexical ambiguity via automatically ordering dictionary definitions or assigning an appropriate meaning to a given context (Rosa & Eskenazi, 2011). In Eom (2012), a captioning tool facilitates listening by providing cues for ambiguous or difficult words, where a word sense disambiguation tool finds suitable definitions for words with multiple meanings.

Third, part of speech (POS) tagging is increasingly needed for language learners for effective word processing. The popularity of POS is mainly because of its ability to provide helpful information (e.g., language morphology, syntax, and phonology) to improve language proficiency (Hamouda, 2013). In an Indonesian computer-assisted self-learning system (Muljono et al., 2017), a POS, tagging with a hidden Markov model, deals with ambiguity by reducing tagging errors in unknown words.

Additionally, NLP also facilitates automatic feedback, i.e., grammar correction and writing evaluation and translation. In Lee et al. (2015), Genie Tutor assists English learning by identifying grammar mistakes and providing correction suggestions. With Genie Tutor, language learners know their mistakes in real-time and learn native expressions. An automatic translation chatbot (Sato et al., 2018) offers different types of second language translation along with first language texts during online interaction. By providing second language input and reducing learners' doubts about their second language competence, the chatbot lowers learners' anxiety and facilitates their language performance and motivation during online collaborations.

4.3. Educational robots for AI education

Educational robots are useful for motivating learners and solidifying abstract and complex topics (e.g., AI education). In Martínez-Tenor et al. (2019), Lego® Mindstorms robots teach reinforcement learning algorithms in a cognitive robotics course. Learners engage in lab exercises by implementing reinforcement learning in coding programs to control real robot movements (e.g., simple wandering, backward/forward motion, and detecting and avoiding obstacles). By converting reinforcement learning theory into real-world problems, learners create their own learning experiences by engaging with both theoretical algorithms and physical implementations. SyRoTek (Kulich et al., 2012) allows remote access to fully autonomous mobile robots placed around reconfigurable obstacles. With SyRoTek, learners control the robots in real-time using self-developed algorithms and then observe how the real robots behave through live videos, thus improving their problem-solving ability by integrating theory into practice.

4.4. EDM for performance prediction

Predicting student performance is important in EDM for mining meaningful patterns and knowledge from large-scale educational data using ML and data mining. EDM's effectiveness in learning to predict has been widely reported. Typical prediction scenarios include academic performance, learner enrolment, dropouts, retention, and early detection of at-risk learners. As for data used for predicting attrition, transcript-based features outperform those based on learner histories prior to college (Aulck et al., 2019). Features derived from institutions' routine data are effective for graduation and re-enrolment prediction. Considering algorithmic performance, Beaulac and Rosenthal (2019) highlight random forests' effectiveness for different prediction tasks, including the prediction of the number of registered learners in future years, learner distribution prediction across programs and at-risk learner identification (e.g., academic failure or dropping out).

4.5. Discourse analysis in CSCL

Collaborative dialogue analysis is essential for facilitating CSCL (Lin & Chan, 2018) as it promotes an understanding of the collaborative process and enables tailored interventions and appropriate scaffolding (Dowell et al., 2019). Jointly using time series analysis and semantic similarity can filter online discourse to identify learners' key collaborative moments (Samoilescu et al., 2019). Informed by the degree of collaboration, which is automatically assessed among learners in their conversations, instructors can provide feedback to promote learner involvement and collaboration in CSCL. Focusing on facilitating large-scale collaborative dialogue data analysis, Shibata et al. (2017) train and test an automatic coding approach based on DL, showing DL's superiority over naive Bayes and support vector machines for supporting authentic learning through monitoring and scaffolding non-activated groups in real-time.

4.6. Neural networks for teaching evaluation

With the rapid growth of higher education, teaching quality has been put in the spotlight. ANNs are revolutionizing teaching quality evaluation by avoiding human subjectivity to enhance evaluation accuracy and effectiveness (Hongmei, 2013). Such neural network-driven models can be further enhanced by particle swarm optimization for weight optimization and modification in accuracy calculation during model training (Rashid & Ahmad, 2016).

4.7. Affective computing for learner emotion detection

Affect in learning is receiving more attention to better understand learner emotions and cognition and to provide affective intervention and support to increase learner self-concept and motivation (Hwang et al., 2020b). Two affective computing techniques (i.e., emotion recognition from physiological or facial expression data and emotion recognition from texts) are widely embodied in ITSs. In Mehmood and Lee (2017), special school instructors teach learners with emotional disorders using wearable sensors and intelligent emotion detection technologies to identify useful information from brain signals. Then, the learners' feelings (i.e., happiness, calm, sadness, and fear) are extracted from the information and processed using support vector machines and near k-neighbor classifiers. In Su et al. (2016), emotions are identified through joint use of facial expression detection and textual sentiment analysis. Such a combined strategy strengthens recognition effectiveness and allows the detection of diverse emotions to facilitate personalized instruction and curriculum content provision.

4.8. Recommender systems for personalized learning

Recommender systems are increasingly integrated into ITSs to generate personalized recommendations about learning resources and paths by considering learners' background knowledge, behavioral preferences, profiles, and interests (Ma & Ye, 2018). In Liu et al. (2018), learners' quiz scores and multi-modal sensing data (i.e., heartbeats, blinks, and facial expressions) are measured to track learning processes and generate personalized guidance based on their present learning states. Such personalized systems can be improved by modifying dynamic key-value memory to design memory structures based on the course's concept list, plus by mapping exercise-concept relations during learners' knowledge tracing (Ai et al., 2019). This helps build learner simulators for exercise recommendation policy training to maximize learners' knowledge level through deep reinforcement learning.

4.9. Challenges and the future of AIEd

This section discusses the challenges existing within the above-discussed themes and points towards future efforts needed to resolve such challenges.

4.9.1. Personalization versus data privacy

The global prevalence of personalized learning calls for more investigations into AI's most effective use to support personalized learning (e.g., adaptively recommending learning materials and scaffolding learners' problem-solving) (Chen et al., 2021). However, to provide personalized experiences, large-scale learner data, which is sometimes highly personal, are required for AI model training. Models sometimes inadvertently store training data with sensitive information that is revealed through model analysis. However, an ML model's potential can only be realized by analyzing learners' data (Chan & Zary, 2019). Since most established models cannot guarantee output models' generalization away from individual learner specifics, plus the uncertainty of data protection places learners' data at risk and lowers AI societal acceptance, there is a need to limit instructors' access to learners' data to meaningfully bound learners' exposure to instructors' knowledge. Educational institutions should be transparent about learner data privacy practices to alleviate data use misperceptions and concerns.

4.9.2. Challenges and ways to increase instructors' AI acceptance

AIEd aims to use AI to facilitate the instruction process (e.g., understand and facilitate CSCL through discourse analysis and achieve performance prediction through EDM), during which instructors are essential, and their acceptance of AI is important. However, as AI is a relatively new concept for instructors, less-experienced instructors usually struggle to execute effective, on-the-spot responses to analytics from AI-empowered applications (Holstein et al., 2017), leading to their reluctance and lower acceptance of AI (Lin et al., 2017). This hinders AI's pedagogical purpose; thus, the improvement of instructors' acceptance of AI systems appears essential.

One way to enhance instructors' confidence in AI is to show the effectiveness of AI systems via robust experiments, particularly under the guidance of time-honored educational theories and philosophies. However, most current AIEd studies fail to positively assess AI system effectiveness through experiments that compare AI's use and traditional instructions (Zawacki-Richter et al., 2019). Albacete et al. (2019) evaluate the effectiveness of Rimac, a natural-language tutoring system capable of dynamically updating learner models, by comparing it with its control version without the updating function. Such an experimental design is challenging due to strict requirements, especially for AI system evaluation, where large samples are required to generate probabilistic results. Additionally, pre- and post-tests are fundamental to objective analysis, and participants should have a similar knowledge level before interventions. Consequently, the effectiveness of AI-driven educational systems is seldom assessed. Nevertheless, such experimental comparisons are indispensable for enhancing instructors' confidence in AI. Researchers should also reach beyond examining how AI improves subject-related outcomes to examining the effectiveness of systems in improving specific abilities (e.g., self-efficacy and higher-order thinking). Thus, in line with Tang et al. (2021), we suggest further investigations into AI's impact on learners' higher-order thinking skills to help deepen instructors' understanding of effective techniques specified for educational goals.

Another approach is to involve instructors in AI system design. Currently, most AI applications/mechanisms remain as proposals, i.e., they are still hypothetical without evidence for their effectiveness in the real world. Hence, real-life decision-support tool development should be promoted to see whether AI-oriented applications can adapt to realistic educational scenarios and be used as pedagogical instruments (Ijaz et al., 2017). However, developing such intelligent systems is complex when learning objectives are considered. Therefore, different types of design and prototyping approaches are desired to allow both data scientists and non-technical stakeholders such as educational experts to be meaningfully involved in system development (Holstein et al., 2018a). Engineers and data scientists are primarily concerned with AI system accuracy in predicting results and less about pedagogical practice. The development of efficient systems specified for particular learning objectives requires connecting closely with pedagogical innovations and carefully considering students' learning styles. Thus, researchers should actively collaborate with subject matter experts or professional educators to build educationally sound AI systems (Chen et al., 2020a). Involving subject matter experts is essential in the AI-building process to steer data scientists in the right direction (Burgess, 2017) to ensure that new models work properly and are applied correctly to whatever dataset is of interest.

Additionally, sufficient technical support is needed to assist instructors in understanding and using AI systems. Instructors are usually challenged by personalized ITSs as they are tasked with monitoring divergent activities simultaneously (Holstein et al., 2018b). Thus, there is a need to examine different types of real-time support offered by AI applications across instructors with varied experience levels. Specifically, researchers should explore how human and automated instruction can most effectively be combined to best support instruction. Such systems have been built on teachers' prior instruction to shape pathways for current instruction and provide guidance on future instruction. These personalized and adaptive AI systems suited to a variety of pedagogical needs are more accessible to instructors (Holstein et al., 2018a), leading to a greater level of personalization across education as a whole by helping instructors design the most effective classroom experience and drive digital transformation.

4.9.3. Shifting from ML to DL

Currently, prevalent techniques in AIEd involve EDM, NLP, discourse analysis, educational robots, ITSs, affective computing, recommender systems, and neural networks, while advanced DL algorithms are less adopted. Considering DL's advantages over traditional ML algorithms in various tasks such as prediction and classification, future studies may show how DL algorithms can replace the ML algorithms already integrated into the existing systems. This would validate DL's effectiveness for multi-task prediction in EDM (e.g., student dropout and use of hints) and reduce implementation time since many required modules already exist (Krouska et al., 2019).

Attention should be paid to DL's generalization ability for adapting or applying it to various new and unexpected tasks. Gray and Perkins (2019) highlight a shortcoming of current ML models' effectiveness for learner outcome prediction because, in many cases, different patterns are often detected for different learner cohorts progressing through courses. Thus, although current models generalize well to test sets, they may not work well for new cases due to implicit memorization of certain examples, leading to ongoing AI model training by constantly including new data and eliminating aging ones. Such processes are repetitive, tedious, and inefficient due to the challenge regarding whether and what attributes and variables within a new dataset should be exploited to improve model performance (Livieris et al., 2019). The following are directions to consider while developing DL-based generalized applications.

There are always new attributes potential to impact AI models' effectiveness that are either currently unavailable but can be collected by instructors or are hidden within students' learning interactions with educational systems (Livieris et al., 2019). There are also features that need constant adjustments, an example of which is the number of days absent indicating potential school-leavers. Thus, future work on automatic feature selection and adjustment is required to facilitate DL model training.

In feature design, we suggest integrating features available in the literature and variables obtained from various channels (e.g., learners' eye-tracking data and electrodermal activities) into modeling to enhance a models' predictive performance (Olive et al., 2019) via feature selection to identify valuable features to predict interested variables. The feature selection can be optimized by considering pedagogical practices and task independence. An example is a pedagogically and theoretically sound feature design assisted by a better understanding of manual grading criteria when developing AI systems for an automatic non-native learner essay assessment. Additionally, it is essential to develop an in-depth understanding of an input feature's relationships and roles to enhance its visibility on learning processes through straightforward visualization and statistical analysis (e.g., structural equation modeling to mediate affective factors' effects).

To initialize the model, most studies train separate classifiers for individuals, which is computationally expensive depending on the dataset, and it also burdens the system. General classifiers trained beforehand and capable of classifying an individual's learning states are needed. Alternative methods include: (1) initializing models with random weights for architecture evaluation with accurate non-linearities and pooling, and (2) exploiting hypernetworks for initialization by inputting learner model architecture and generating model weights. The latter strategy also reduces the learners' burden on model training and promotes the learners' perceived ease of use without requiring them to report learning states for classifier training.

Additionally, overfitting should be avoided and over-sampling impact reduced by testing a model's effectiveness in various scenarios, including: (1) experiments on large sample sizes, (2) applying it in different contexts (e.g., blended learning), courses, and institutions (e.g., middle school and college students), and (3) considering learners' demographic characteristics (e.g., gender, culture, and high/low performance) to validate a models' general effectiveness.

5. Conclusion

This first-in-depth topic-based bibliometric study tracks current advances in AIEd research in the first two decades of the 21st century, which is needed as AIEd is receiving increasing attention. Methodologically, bibliometric indicators such as the H-index and publication count measuring scientific impact and productivity help identify active sources, countries/regions, and institutions in AIEd research. This enables scholars to be more aware of channels to make contributions and important actors to learn from (Chen et al., 2020b). Social network analysis, through scientific collaboration visualization, also reveals an invisible collaborative network of participating countries/regions and institutions in AIEd research, intuitively helping to show collaborative relationships and potential scientific collaborators (Chen et al., 2018). Additionally, topic modeling, capable of mining themes from large-scale textual data, helps understand the past and present AIEd scientific structure (Roberts et al., 2014). The identified topics and themes were further analyzed using the MK test to reveal topic dynamics to indicate how research foci change and develop, providing insights into AIEd's future directions (Chen et al., 2020a). Increasingly diverse AI technologies are being incorporated into various applications (e.g., ITSs, robots, mobile devices, and digital games) to facilitate teaching and learning. Analytical techniques such as ML, EDM, NLP, ANNs, and affective computing are commonly adopted for analyzing large-scale data from various educational scenarios (e.g., computer science education, language education, STEM education, special education, virtual surgery training, CSCL, and flipped classrooms). Eight promising areas within AIEd include (1) ITSs for special education; (2) NLP for language education; (3) educational robots for AI education; (4) EDM for performance prediction; (5) discourse analysis in CSCL; (6) neural networks for teaching evaluation; (7) affective computing for learner emotion detection; and (8) recommender systems for personalized learning. Finally, we also highlight the need to: (1) be transparent about learner data usage to realize personalized learning, (2) enhance instructors' AI acceptance by involving them in system design and convincing them of AI's effectiveness through robust experimental design, and (3) move towards "DLEd" for educational system design with higher generalizability.

Acknowledgement

An abstract entitled "A Bibliometric Study on Artificial Intelligence in Education for Two Decades" from this paper was presented at the International Conference on Education and Artificial Intelligence 2020, The Education University of Hong Kong, 9-11 November 2020, Hong Kong. Haoran Xie's work in this research is supported by the Direct Grant (DR21A5) and the Faculty Research Fund (DB21A4) at Lingnan University, Hong Kong. Gary Cheng's work is supported by General Research Fund (No. 18601118) of Research Grants Council of Hong Kong SAR, China, and One-off Special Fund from Central and Faculty Fund in Support of Research from 2019/20 to 2021/22 (MIT02/19-20) of The Education University of Hong Kong.

References

Ai, F., Chen, Y., Guo, Y., Zhao, Y., Wang, Z., & Fu, G. (2019). Concept-aware deep knowledge tracing and exercise recommendation in an online learning system. *International Conference on Educational Data Mining (EDM)* (pp. 240–245). Montreal, Canada: International Educational Data Mining Society. Retrieved from https://files.eric.ed.gov/fulltext/ED599194.pdf

Albacete, P., Jordan, P., Katz, S., Chounta, I.-A., & McLaren, B. M. (2019). The Impact of student model updates on contingent scaffolding in a natural-language tutoring system. In *International Conference on Artificial Intelligence in Education* (pp. 37–47). https://doi.org/10.1007/978-3-030-23204-7_4

Aulck, L., Nambi, D., Velagapudi, N., Blumenstock, J., & West, J. (2019). Mining university registrar records to predict first-year undergraduate attrition. *International Conference on Educational Data Mining (EDM)* (pp. 9–18). International Educational Data Mining Society. https://files.eric.ed.gov/fulltext/ED599235.pdf

Bayne, S. (2015). Teacherbot: Interventions in automated teaching. *Teaching in Higher Education*, *20*(4), 455–467. https://doi.org/10.1080/13562517.2015.1020783

Beaulac, C., & Rosenthal, J. S. (2019). Predicting university Students' academic success and major using random forests. *Research in Higher Education*, *60*(7), 1048–1064. https://doi.org/10.1007/s11162-019-09546-y

Burgess, A. (2017). *The Executive guide to artificial intelligence: How to identify and implement applications for AI in your organization*. Springer. https://doi.org/10.1007/978-3-319-63820-1

Chan, K. S., & Zary, N. (2019). Applications and challenges of implementing artificial intelligence in medical education: Integrative review. *JMIR Medical Education*, *5*(1), e13930. https://doi.org/10.2196/13930

Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial Intelligence trends in education: A Narrative overview. *Procedia Computer Science*, *136*, 16–24. https://doi.org/10.1016/j.procs.2018.08.233

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*, 1–14. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1016/j.caeai.2020.100005

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. https://www.jstor.org/stable/10.2307/26977868

Dowell, N., Lin, Y., Godfrey, A., & Brooks, C. (2019). Promoting inclusivity through time-dynamic discourse analysis in digitally-mediated collaborative learning. In *International Conference on Artificial Intelligence in Education* (pp. 207–219). https://doi.org/10.1007/978-3-030-23204-7_18

Eom, S. (2012). *Automatic presentation of sense-specific lexical information in an intelligent learning system* (Order No. 3524085) [Doctoral dissertation, Georgetown University]. ProQuest Dissertations and Theses Global.

Goksel, N., & Bozkurt, A. (2019). Artificial Intelligence in Education: Current insights and future perspectives. In *Handbook of Research on Learning in the Age of Transhumanism* (pp. 224–236). IGI Global. https://doi.org/10.4018/978-1-5225-8431-5.ch014

Gray, C. C., & Perkins, D. (2019). Utilizing early engagement and machine learning to predict student outcomes. *Computers & Education*, *131*, 22–32. https://doi.org/10.1016/j.compedu.2018.12.006

Guan, C., Mou, J., & Jiang, Z. (2020). Artificial intelligence innovation in education: A Twenty-year data-driven historical analysis. *International Journal of Innovation Studies*, *4*(4), 134–147. https://doi.org/10.1016/j.ijis.2020.09.001

Hamouda, A. (2013). A Study of dictionary use by Saudi EFL students at Qassim University. *Study in English Language Teaching*, *1*(1), 228–257. https://doi.org/10.22158/selt.v1n1p227

Hinojo-Lucena, F.-J., Aznar-Díaz, I., Cáceres-Reche, M.-P., & Romero-Rodríguez, J.-M. (2019). Artificial intelligence in higher education: A Bibliometric study on its impact in the scientific literature. *Education Sciences*, *9*(1), 51. https://doi.org/10.3390/educsci9010051

Holstein, K., Hong, G., Tegene, M., McLaren, B. M., & Aleven, V. (2018b). The Classroom as a dashboard: Co-designing wearable cognitive augmentation for K-12 teachers. *The 8th international conference on learning Analytics and knowledge* (pp. 79–88). https://doi.org/10.1145/3170358.3170377

Holstein, K., McLaren, B. M., & Aleven, V. (2017). SPACLE: Investigating learning across virtual and physical spaces using spatial replays. *The Seventh International Learning Analytics & Knowledge Conference* (pp. 358–367). https://doi.org/10.1145/3027385.3027450

Holstein, K., McLaren, B. M., & Aleven, V. (2018a). Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. *International Conference on Artificial Intelligence in Education* (pp. 154–168). https://doi.org/10.1007/978-3-319-93843-1_12

Hongmei, L. (2013). Application research of BP neural network in English teaching evaluation. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, *11*(8), 4602–4608. http://dx.doi.org/10.11591/telkomnika.v11i8.3085

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020a). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, *1*, 100001. https://doi.org/10.1016/j.caeai.2020.100001

Hwang, G. J., Sung, H. Y., Chang, S. C., & Huang, X. C. (2020b). 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*, 100003. https://doi.org/10.1016/j.caeai.2020.100003

Ijaz, K., Bogdanovych, A., & Trescak, T. (2017). Virtual worlds vs books and videos in history education. *Interactive Learning Environments*, *25*(7), 904–929. https://doi.org/10.1080/10494820.2016.1225099

Krouska, A., Troussas, C., & Virvou, M. (2019). SN-Learning: An Exploratory study beyond e-learning and evaluation of its applications using EV-SNL framework. *Journal of Computer Assisted Learning*, *35*(2), 168–177. https://doi.org/10.1111/jcal.12330

Kulich, M., Chudoba, J., Kosnar, K., Krajnik, T., Faigl, J., & Preucil, L. (2012). SyRoTek—Distance teaching of mobile robotics. *IEEE Transactions on Education*, *56*(1), 18–23. https://doi.org/10.1109/TE.2012.2224867

Lee, K., Kwon, O.-W., Kim, Y.-K., & Lee, Y. (2015). A Hybrid approach for correcting grammatical errors. In F. Helm, L. Bradley, M. Guarda, & S. Thouesny (Eds), *Critical CALL–Proceedings of the 2015 EUROCALL Conference, Padova, Italy* (pp. 362–367). Dublin, Ireland: Research-publishing.net. http://dx.doi.org/10.14705/rpnet.2015.000359

Lin, F., & Chan, C. K. K. (2018). Examining the role of computer-supported knowledge-building discourse in epistemic and conceptual understanding. *Journal of Computer Assisted Learning*, *34*(5), 567–579. https://doi.org/10.1111/jcal.12261

Lin, C.-C., Liu, G.-Z., & Wang, T.-I. (2017). Development and usability test of an e-learning tool for engineering graduates to develop academic writing in English: A Case study. *Educational Technology & Society*, *20*(4), 148–161.

Liu, S., Chen, Y., Huang, H., Xiao, L., & Hei, X. (2018). Towards smart educational recommendations with reinforcement learning in classroom. In *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)* (pp. 1079–1084). https://doi.org/10.1109/TALE.2018.8615217

Livieris, I. E., Drakopoulou, K., Tampakas, V. T., Mikropoulos, T. A., & Pintelas, P. (2019). Predicting secondary school students' performance utilizing a semi-supervised learning approach. *Journal of Educational Computing Research*, *57*(2), 448–470. https://doi.org/10.1177/0735633117752614

Ma, X., & Ye, L. (2018). Career goal-based e-learning recommendation using enhanced collaborative filtering and prefixspan. *International Journal of Mobile and Blended Learning (IJMBL)*, *10*(3), 23–37. https://doi.org/10.4018/IJMBL.2018070103

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica: Journal of the Econometric Society*, *13*(3), 245–259. https://doi.org/10.2307/1907187

Martínez-Tenor, Á., Cruz-Martín, A., & Fernández-Madrigal, J.-A. (2019). Teaching machine learning in robotics interactively: the case of reinforcement learning with Lego® Mindstorms. *Interactive Learning Environments*, *27*(3), 293–306. https://doi.org/10.1080/10494820.2018.1525411

Mehmood, R. M., & Lee, H. J. (2017). Towards building a computer aided education system for special students using wearable sensor technologies. *Sensors*, *17*(2), 317. https://doi.org/10.3390/s17020317

Mondragon, A. L., Nkambou, R., & Poirier, P. (2015). Towards an Integrated Specialized Learning Application (ISLA) to support high functioning ASD children in mathematics learning. In *European Conference on Technology Enhanced Learning* (pp. 225–239). https://doi.org/10.1007/978-3-319-24258-3_17

Mondragon, A. L., Nkambou, R., & Poirier, P. (2016). Towards an effective affective tutoring agent in specialized education. In *International Conference on Intelligent Tutoring Systems* (pp. 402–408). https://doi.org/10.1007/978-3-319-39583-8_48

Muljono, M., Afini, U., Supriyanto, C., & Nugroho, R. A. (2017). The Development of Indonesian POS tagging system for computer-aided independent language learning. *International Journal of Emerging Technologies in Learning (IJET)*, *12*(11), 138–150. https://doi.org/10.3991/ijet.v12i11.7383

Olive, D. M., Huynh, D. Q., Reynolds, M., Dougiamas, M., & Wiese, D. (2019). A Quest for a one-size-fits-all neural network: Early prediction of students at risk in online courses. *IEEE Transactions on Learning Technologies*, *12*(2), 171–183. https://doi.org/10.1109/TLT.2019.2911068

Panaite, M., Dascalu, M., Dessus, P., Bianco, M., & Trausan-Matu, S. (2018). Identifying reading strategies employed by learners within their oral French self-explanations. *The 5th International Workshop on "Semantic and Collaborative Technologies for the Web", joint to the 14th International Scientific Confrerence on eLearning and Software for Education (eLSE 2018)* (pp. 362–368). Advanced Distributed Learning Association. https://doi.org/10.12753/2066-026X-18-120

Rashid, T. A., & Ahmad, H. A. (2016). Lecturer performance system using neural network with Particle Swarm Optimization. *Computer Applications in Engineering Education*, *24*(4), 629–638. https://doi.org/10.1002/cae.21737

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R package for structural topic models. *Journal of Statistical Software*, *91*(2), 1-40. https://doi.org/10.18637/jss.v091.i02

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. https://doi.org/10.1111/ajps.12103

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, *26*(2), 582–599. https://doi.org/10.1007/s40593-016-0110-3

Rosa, K. Dela, & Eskenazi, M. (2011). Impact of word sense disambiguation on ordering dictionary definitions in vocabulary learning tutors. *The Twenty-Fourth International Florida Artificial Intelligence Research Society Conference* (pp. 507–512). Palm Beach, FL: AAAI Press. https://www.aaai.org/ocs/index.php/FLAIRS/FLAIRS11/paper/view/2545/3096

Samoilescu, R.-F., Dascalu, M., Sirbu, M.-D., Trausan-Matu, S., & Crossley, S. A. (2019). Modeling collaboration in online conversations using time series analysis and dialogism. *International Conference on Artificial Intelligence in Education* (pp. 458–468). https://doi.org/10.1007/978-3-030-23204-7_38

Sato, T., Ogura, M., Aota, S., & Burden, T. (2018). Examining the impact of an automated translation chatbot on online collaborative dialog for incidental L2 learning. In P. Taalas, J. Jalkanen, L. Bradley & S. Thouesny (Eds.), *Future-Proof CALL: Language Learning as Exploration and Encounters–Short Papers from EUROCALL 2018* (pp. 284–289). http://doi.org/10.14705/rpnet.2018.26.851

Shibata, C., Ando, K., & Inaba, T. (2017) Towards automatic coding of collaborative learning data with deep learning technology. *eLmL 2017: The Ninth International Conference on Mobile, Hybrid, and On-line Learning* (pp. 65–71). Nice, France: IARIA. https://laweb.cloud.teu.ac.jp/cms/wp-content/uploads/2017/08/contribution_50092_eLmL_201703032017ando2.pdf

Su, S.-H., Lin, H.-C. K., Wang, C.-H., & Huang, Z.-C. (2016). Multi-modal affective computing technology design the interaction between computers and human of intelligent tutoring systems. *International Journal of Online Pedagogy and Course Design (IJOPCD)*, *6*(1), 13–28. http://doi.org/10.4018/IJOPCD.2016010102

Sun, Q., & Winoto, P. (2019). An Intelligent LEGO tutoring system for children with autism spectrum disorder. *The 2019 the International Conference on Pattern Recognition and Artificial Intelligence* (pp. 56–59). https://doi.org/10.1145/3357777.3357795

Tang, K. Y., Chang, C. Y., & Hwang, G. J. (2021). Trends in artificial intelligence-supported e-learning: A Systematic review and co-citation network analysis (1998–2019). *Interactive Learning Environments*, 1–19. https://doi.org/10.1080/10494820.2021.1875001

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. https://doi.org/10.1186/s41239-020-00186-2

Zawacki-Richter, O., Marín, V. I., 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. https://doi.org/10.1186/s41239-019-0171-0

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. https://doi.org/10.1080/09588221.2020.1744666

Li, J., Liu, J., Yuan, R., & Shadiev, R. (2022). The Influence of Socially Shared Regulation on Computational Thinking Performance in Cooperative Learning. *Educational Technology & Society*, *25* (1), 48-60.

The Influence of Socially Shared Regulation on Computational Thinking Performance in Cooperative Learning

Jiansheng Li, Jiao Liu, Rui Yuan and Rustam Shadiev*
School of Education Science, Nanjing Normal University, Nanjing, China // 2869753244@qq.com // 2776571510@qq.com // 1622529878@qq.com // rustamsh@gmail.com
*Corresponding author

(Submitted March 9, 2021; Revised July 13, 2021; Accepted July 26, 2021)

ABSTRACT: This study explores the role of socially shared regulation on computational thinking performance in cooperative learning. Ninety-four middle school students from China aged between 16 and 18 participated in this study. Forty-six students were in the experimental group, and 48 students were in the control group. Students in the experimental group learned under the socially shared regulation of learning (SSRL) condition, which included planning and goal setting, task and content monitoring, and task and content evaluation. Students in the control group learned in a traditional way. The results showed that the students in the experimental group significantly outperformed their counterparts on the midtest and posttest. Additionally, the learning gain of the experimental group was much better from the pretest to the midtest. Different subgroups in the experimental group had different learning performances, and task monitoring and content monitoring were two important SSRL processes that led to improved computational thinking performance. Our results suggest that SSRL is beneficial for learning computational thinking subjects. Throughout the process of SSRL, different groups have different learning dynamics, and task and content monitoring plays a major role in computational thinking performance.

Keywords: Socially shared regulation, Group monitoring, Cooperation, Computational thinking, Learning performance

1. Introduction

With the development of artificial intelligence, cultivating students' computational thinking has become an important matter (Angeli et al., 2016). Related research explores how computational thinking in programming can be cultivated not only on the individual level but also through nurturing such skills through cooperative learning activities (Shadiev et al., 2014). For example, McDowell et al. (2003) carried out an experimental study on pair programming and found that pair programming was better for improving students' learning outcomes (i.e., learning performance, perception and persistence) than individual programming. Ge (2014) proposed the collaborative learning model based on computational thinking for training computational thinking and applied it to middle-school classroom teaching. Those results were positive and demonstrated the positive effects of the model on computational thinking skills. Denner et al. (2014) carried out research on computational thinking through cooperative task completion and obtained similar results, i.e., cooperative learning was beneficial for learning outcome improvement, especially for students who had less programming experience.

How learning interactions occur in cooperative learning and how they affect performance are important questions in the field and ones that scholars have attempted to solve in their research. Chi et al. (2014) and Hwang et al. (2012) explained that interaction in cooperative learning occurs when students share and present different perspectives on the problem-solving process. Hadwin et al. (2017) argued that social coregulation is an important component of successful cooperation. For example, individuals in the group negotiate a common goal, adjust the goal according to their own abilities, and discuss common goals and group progress. Without the skills and willingness of individuals and groups to act cooperatively, learning interactions cannot occur (Hadwin & Oshige, 2011b; Hwang et al., 2015). Furthermore, Hadwin et al. (2011a) emphasized the monitoring process by students in learning, as it is particularly important in the regulation process. However, little is known about how group regulation occurs in cooperative learning, whether it exerts an impact on computational thinking performance, or what the main influencing factors are, and there remain many other similar unanswered questions in the field. This paper thus attempts to answer them.

2. Background of study

2.1. Computational thinking and cooperative learning

Computational thinking involves solving problems, designing systems, and understanding human behavior by drawing on the concepts fundamental to computer science (Wing, 2006). Computational thinking is reformulating a seemingly difficult problem into one we know how to solve, perhaps by reducing, embedding, transforming, or simulating (Angeli et al., 2016). In large, complex tasks or designing a large, complex system, abstraction and decomposition are often used to model relevant aspects of the problem and make it easier to deal with by choosing the appropriate representation for the problem. Such an approach is planning, learning, and scheduling in the presence of uncertainty. Computational thinking is thinking in terms of prevention, protection, and recovery from worst-case scenarios through redundancy, damage containment, and error correction (Wing, 2006).

Aho (2012) considered computational thinking to be the thought processes involved in formulating problems such that their solutions can be represented as computational steps and algorithms. Computational thinking includes (but is not limited to) the following steps (International Society for Technology in Education, 2021): (a) formulating problems in a way that enables the use of a computer and other tools to help solve them; (b) logically organizing and analyzing data; (c) representing data through abstractions such as models and simulations; (d) automating solutions through algorithmic thinking (a series of ordered steps); and (e) identifying, analyzing, and implementing possible solutions with the goal of achieving the most efficient and effective combination of steps and resources. Therefore, computational thinking can be cultivated through problem calculation steps and algorithms. For example, Brennan and Resnick (2012) claimed that computational thinking cultivation involves the following: (1) computational thinking concepts: sequences, loops, events, parallelism, conditionals, operators, data; (2) computational thinking practices: being incremental and iterative, testing and debugging, reusing and remixing, abstracting and modularizing; and (3) computational thinking perspectives: expressing, connecting and questioning. Similarly, according to Barefootcomputing (2021), computational thinking cultivation involves two aspects: concepts (logic, algorithms, decomposition, patterns, abstraction, evaluation) and approaches (tinkering, creating, debugging, persevering, collaborating).

The simplest measure of the results of computational thinking cultivation is the computational thinking test. The goal of the computational thinking test is to assess students' ability to solve complex problems using computational thinking by asking students to solve practical problems. For example, the Bebras Tasks, which most researchers have used in computational thinking (Dagienė & Futschek, 2008), have been noted as more than likely to be a foundation for a future PISA (program for international student assessment) test in the field of computer science (Román-González et al., 2019; Yağcı, 2019).

Cooperative learning is also concerned with the cultivation of students' computational thinking. Mcdowell et al. (2003) examined the effectiveness of pair programming in programming courses and found that students who used pair programming produced better programs and were more confident in their problem solving. This result was also verified in other studies (e.g., in Denner et al., 2014). Similarly, Turchi et al. (2019) believed that cooperative game-based learning could foster students' computational thinking skills.

2.2. Socially shared regulation of learning

There are three regulation modes in collaborative learning: self-regulation, coregulation and socially shared regulation (Winne & Hadwin, 1998). Self-regulation refers to the individual's regulation of cognition, metacognition, motivation, emotion and behavior to adapt to other members of the group (Hu & Driscoll, 2013). Coregulation emphasizes the influence among individuals, which means that learners adjust their learning strategies when they interact with other members of the group (Zheng et al., 2017). Socially shared regulation emphasizes conscious, strategic and interactive planning, task formulation, reflection and adaptation within the group (Winner et al., 2011; Hadwin et al., 2011a). In the educational context, such modes are called self-regulated learning, coregulated learning and socially shared regulation of learning (SSRL).

Of particular interest to this study is SSRL, which occurs in the learning process when group members complement one another's cognitive resources; that is, they set common goals together, share responsibility for appropriate strategy to formulate goals, and coordinate changes and adjustments to optimize the problem-solving process (Miller et al., 2017). SSRL seems best to mirror egalitarian, complementary monitoring and regulation over the task, thus bringing the research closer to phenomena relevant to joint, peer-mediated learning. SSRL is committed to common regulatory activities (Vauras et al., 2003). Therefore, SSRL involves different aspects of regulation to ensure that group members remain involved and provide consistent efforts. Low-level social regulation involves the simple exchange or sharing of facts and clarification of understanding, while high-level social regulation is characterized by the use of both shared regulation and deep-level content processing. In SSRL, group members actively and cooperatively monitor developed ideas (Rogat et al., 2011). They also regulate their metacognition, cognition, motivation and behavior (Hadwin et al., 2011a). Moreover, students share multiple ideas and perspectives to be weighed and negotiated (Järvelä et al., 2013). In essence, the SSRL begins to expand regulation activities to include the negotiation and regulation of the group's collective activities. Lee (2014) identified several socially shared regulation processes in computer-supported collaborative learning: planning, goal setting, task monitoring, content monitoring, task evaluation, and content evaluation (Table 1). At the group cooperation discourse level, Volet et al. (2013) believed that all regulation activities in the group come from questions (direct or implicit questions), explanatory statements or abstracts (usually tentative) or implicit or triggered suggestions (under certain background knowledge).

Table 1. Socially shared regulation process

Code	Definition
Planning and Goal Setting	• Presenting a question as a starter for the group's plan or goal • Posting the guiding questions as a starter • Discussing plans and goals • Expressing agreement
Task Monitoring	• Verifying the progress or the completion of each guiding question • Checking the time • Correcting typos
Content Monitoring	• Providing a reason to support the responses or ideas • Checking the accuracy of the task responses
Task Evaluation	• Checking the completion of all the guiding questions
Content Evaluation	• Checking whether the group met its initial goals • Checking whether the group's views were in agreement • Evaluating whether the group completed the task

From what has been discussed above, this study uses the group cooperation process (Lee, 2014), group cooperation discourse level (Volet et al., 2013), and computational thinking definition (Brennan & Resnick, 2012) to analyze the discourse of group cooperation and to explore group regulation activities and their role. Therefore, the following research questions will be addressed in this paper:

- Does SSRL affect computational thinking performance?
- In SSRL activities, which process leads to the improvement of computational thinking performance?
- Do different subgroup dynamics in SSRL activities exert different impacts on computational thinking performance?

3. Method

3.1. Participants

The participants were 94 middle school students aged between 16 years and 18 years from a senior high school in China. They all were at the same learning grade. The students were assigned to an experimental group (n = 46) and a control group (n = 48). There were 29 boys and 17 girls in the experimental group and 24 boys and 24 girls in the control group. All participants were informed about the study and gave informed consent prior to participation in the study.

3.2. Procedure

The teaching experiment lasted for eight weeks and involved the following related aspects: data, sequences, conditionals, loops, abstracting and modularizing, testing and debugging, reusing and recreating. The teaching content of the experimental group was to help students learn basic knowledge of Python. Teaching experiments were conducted in a classroom environment, and group collaborative learning and discussion were presented through Shimo Docs. Shimo Docs (https://shimo.im/) is enterprise office service software that supports clouds and real-time collaboration. It enables multiple users to edit the same document and to have real-time discussion among learners. This software is widely used in educational institutions in China.

The experimental and control groups were taught by the instructor according to the teaching plan. During classroom teaching, textual discussion data in Shimo Docs were collected. Then, the students in each group were divided into 8 subgroups, with 5 or 6 students in each subgroup according to the pretest results and the grouping principle of heterogeneity in the same group and homogeneity in different groups. After that, each subgroup established a discussion area in the platform and named it with their own student number. We designed guidance questions for the groups to facilitate their SSRL, as shown in Table 2.

Table 2. Description of SSRL implementation process

Code	Guidance questions to the group
Planning and Goal Setting	• Asking questions about Python drawing as a starting point for the group's plan or goal. • Issuing the guidance question regarding painting and using it as the starting point for the issue. • Discussing the planning ideas for the questions to be solved.
Task Monitoring	• Verifying the progress or the completion of each guiding question of the group: What have we completed? How much more? • Checking the time: How much time does the group require to complete the drawing? • Correcting spelling mistakes: Checking the spelling of Python code.
Content Monitoring	• Do you think this answer is correct? • Do you agree/disagree with your partner's answer? Everyone must express their own views and give reasons for their ideas. They cannot simply agree with each other. Finally, they have to reach an agreement.
Task Evaluation	• Whether each group completed all the guidance questions about Python painting.
Content Evaluation	• Checking whether the groups have completed the painting pattern we initially imagined. • Checking the use of relevant concepts in all relevant works. Are these drawings related to the knowledge of (loops, conditionals, sequences) in Python? • Evaluating the content to answer the task.

At the beginning of the cooperative learning stage, we required the students in the experimental group to become familiar with the rules of group discussion, and at the end of the learning stage, we asked them to check whether their opinions were aligned. As a result, the students in each group were asked to discuss and negotiate the guidance questions in the cooperation plan and implementation table. The guidance questions in collaboration involved understanding shared tasks (e.g., to describe group learning tasks and the purpose of this task) and shared goals (e.g., to set a goal for a group task) and included common tasks and content monitoring (e.g., Is this answer correct? How well does the task match the instruction? Please explain your responses), and task and content evaluation (e.g., Have we all completed the guidance? Have we achieved our original goal?).

The following is a group content monitoring example. The content monitoring events observed in the SSRL group included monitoring content contribution and understanding, checking for evolving task responses, and monitoring the development of the summary (see Table 2). The first feature of content monitoring was that the group participated in the monitoring process equally. Everyone had to actively elaborate, ask questions and pay attention to the contribution quality of the group members and the task response they negotiated. Next, according to the task requirements, they had to check whether others had completed assignments correctly. In response, all team members were engaged in content monitoring. The following fragment is the students' answer to guidance question 3 in the study.

Guidance question 3: Which statements (loops, conditionals, sequences) should be used to solve the problem? What is the problem?

ID2: Loops, conditionals and sequences can be used.
ID6: Does this require four times loops commands?
ID42: That pentagram has five sides. It should loop five times, shouldn't it?
ID26: Yes, I think so, too.
ID19: I think it's better to loop five times or one time.

ID2: Why is it called a loop? Why loop once?
ID6: Yes, why? Who can tell me why?
ID30: All right, because when there are 5 loops, only one line can be put in the statement block. When there is 1 loop, we can put all five lines in the statement block.
ID19: Yes, that's what I mean.
ID26: What kind of looping is it? "For" or "While"?
ID30: I think it's OK to use "For" or "While" because we can achieve our goal regardless of which statement we use.
ID2: The teacher often uses the "For" loops in the group, so why not use the "While" loops?
ID6: If there is a fixed number of times, we usually use the "For" loops.
ID42: I remember the teacher said this.
ID19: You are so excellent!

Although both groups covered the same learning content in the study, the control group's cooperative learning had no monitoring requirements.

3.3. Data acquisition

First, students were tested on their computational thinking through pretest questions. Second, SSRL was divided into four processes, and the number of regulations in different processes was recorded (see Table 2). Third, a Python knowledge test was conducted in the middle of the study. Fourth, the students' computational thinking was tested again at the end of the study, i.e., posttest. The computational thinking test and a Python knowledge test both included objective questions with objective answers. The experiment was carried out eight times, and the students collaborated on the Shimo Doc in each session. Thus, a total of eight SSRL cooperative learning sessions were conducted. Each class lasted for 45 minutes. We collected SSRL behavior through Shimo Docs. The regulation behavior of students in the SSRL process was based on statistical time and objective evidence.

3.4. The test of computational thinking

The pretest and posttest items were developed based on the Bebras tasks. According to Dagienė and Futschek (2008), the Bebras tasks comprise a set of activities designed within the context of the Bebras International Contest, which is a motivation competition in informatics and computer literacy for students of the lower, middle and upper levels of secondary school. Scholars argue that the Bebras tasks are a valid assessment tool and reliably measure CT skills, especially those that need to be transferred and projected to solve "real-life" problems. For this reason, the Bebras tasks have been widely used by scholars in the educational context (Román-González et al., 2019; Yağcı, 2019). The test of this study consisted of 20 multiple-choice questions. The test was divided into two parts: calculation concept and calculation practice. The calculation concept part tested how well students understood the basic concepts in the course and involved sequences (three questions), loops (two questions), events and parallelism (two questions), conditionals (three questions), operators (two questions) and data (two questions). The calculation practice part sought to test how well students understood practical operations in the course and involved incremental and iterative (two questions), testing and debugging (two questions), reusing and recreating (one question), and abstracting and modularizing (one question) items. A few examples of the tests are presented in the Appendix.

3.5. The test of Python knowledge

A midtest was also carried out in the middle of the experiment. Python was chosen as a programming tool in the study because of its interactive environment and its convenience in allowing beginner programmers to write meaningful but nontrivial programs within a short time (Maria & Tsiatsos, 2017). The test questions were developed by experienced grade instructors based on the knowledge objectives of the course and piloted with a few groups of students beforehand. The test items were of medium difficulty and met the Chinese middle-school information technology curriculum requirements. The test items involved knowledge of the calculation concept and calculation practice, including five fill-in-the-blank questions, ten multiple-choice questions, one programming question (open, the answer was not unique) and one short-answer question.

3.6. The coding framework of SSRL monitoring

Two kinds of monitoring – one in which each group member is responsible for regulating his/her own learning (SRL monitoring) and one in which group members regulate the learning process together (SSRL monitoring) – play an important role in the process of group cooperation (Zheng et al., 2019). In this experiment, group cooperative learning was arranged based on SSRL. Table 3 presents the coding framework for coding online texts.

In Table 2, SSRL can be distinguished by the regulation of object operations, the operations being performed (Malmberg et al., 2017), and activities that include SSRL planning, SSRL tasks, SSRL statements and progress. Students engage in SSRL tasks and planning if they activate their personal knowledge and consider personal behavior. The SSRL task and SSRL planning were represented by statements centered on group tasks and group actions, while SSRL progress involved the time management, conflict resolution and mutual understanding that guided the whole team's efforts.

Table 3. The coding framework of SRL and SSRL (Zheng et al., 2019)

Monitoring	Category	Description	Examples[1]	Features[2]
SRL monitoring	SRL task	Reviewing the prior knowledge required for the task	Well, I need to use the loops	I am… I need…
	SRL planning	Considering personal behavior	Let me see the program I wrote again	Let… I…I think…
	SRL statement	Putting forward their own view of the task		
	SRL progress	The current progress of one person and the whole team		
SSRL monitoring	SSRL planning	Plan setting and setting the purpose of the team	What's the target of our group? What's the purpose?	Let's… We need… Who?
	SSRL task	The next action to complete the task, form a statement or set of statements for other team members		Why? Please tell me…
	SSRL statement	Elaborating ideas and making your reasoning work for the team. Do you agree with others? Why? What's the reason?	I don't agree with you because conditionals make the procedure simple.	
	SSRL progress	Have you achieved your original goal?	Yes	

Note. [1]Examples and features were derived from content created by the participants based on recommendations of Zheng et al. (2019). [2]Features of regulatory activities at the individual or group level.

4. Results

4.1. Does SSRL affect computational thinking performance?

First, we explored whether the two groups were equal in terms of their basic demographic characteristics and prior knowledge. The results of *t*-tests and chi-square analysis demonstrated that the two groups did not differ significantly ($p > .05$) in average age or the proportion of boys and girls across both groups. The first column of Table 4 shows the mean and standard deviation values of the two groups for the tests. According to *t*-test analyses ($t = 1.29$, $p = .20$), the two groups were not significantly different in their pretest scores. Therefore, we concluded that the groups were equivalent in basic demographic characteristics and prior knowledge.

Next, we explored whether the groups differed in computational thinking performance. The second row of Table 4 shows the mean and standard deviation values for the midtest for the two groups. The average score for the experimental group was 62.63, while that for the control group was 44.06, and this difference was significant according to the *t*-test ($t = 4.27$, $p = .00$). To explore the impact of pretest scores on midtest scores, we conducted an analysis of covariance (ANCOVA) with pretest scores as a covariate. ANCOVA results showed that students in the experimental group scored significantly higher than students in the control group on the midtest, $F(1,91) = 4.937$, MSE = 316.210, $p = .029$, $d = 0.308$.

Table 4. Test performance for the two groups

Test	Experimental group		Control group	
	M	*SD*	*M*	*SD*
Pretest	52.97	10.80	49.78	9.78
Midtest	62.63	18.99	44.06	17.06*
Posttest	65.79	26.52	51.56	18.98*

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

The third row of Table 4 shows the mean and standard deviation values on the posttest for the two groups. In the posttest, the computational thinking scores of the experimental group and the control group were 65.79 and 51.56, respectively, and this difference was significant according to the *t*-test ($t = 2.61$, $p = .01$). Similarly, to determine whether the pretest and midtest scores affected the posttest score, we conducted two ANCOVA tests. ANCOVA with pretest scores as a covariate showed that students in the experimental group scored significantly higher than students in the control group on the midtest, $F(1,91) = 6.658$, MSE = 473.237, $p = .011$, $d = 1.023$. However, ANCOVA with midtest scores as a covariate revealed that the experimental and control groups did not perform significantly differently from one another on the posttest, $F(1,91) = 3.816$, MSE = 425.513, $p = .54$, $d = .608$.

Thus, the major empirical finding in this study is that the students in the experimental group had higher learning gains than the students in the control group from the beginning of the study to the middle. However, the learning gain from the middle to the end of the study was not as large as that from the beginning to the middle. This finding suggests that the effect of SSRL on learning gains mainly occurs from the beginning to the middle of a course. That is, the regulation of SSRL was more effective during the first half of the experiment. Therefore, in the second half of the experiment, the regulatory effect was not as obvious.

4.2. In SSRL activities, which process leads to the improvement of computational thinking performance?

Based on Table 2, SSRL includes five processes: planning and goal setting, task monitoring, content monitoring, task evaluation, and content evaluation. In the experimental process, we combined the task evaluation with the content evaluation, and for this reason, we had four SSRL processes in the study. In these four different processes, students' monitoring time was tallied and then analyzed. Analysis of variance (ANOVA) results for experimental students' computational thinking posttest performances with respect to the four SSRL processes are shown in Table 5. According to the results, there were significant improvements in the task and content monitoring processes, $p < .05$. However, no statistically significant results were obtained for the planning and goal setting and the evaluation monitoring processes. Therefore, in the process of SSRL, monitoring (i.e., task and content monitoring) plays a major role in the learning performance.

Table 5. ANOVA results for learning performance with respect to four SSRL processes

Group	Process	*F*	*Sig.*
Experimental group	Planning and Goal Setting	2.395	.081
	Task Monitoring	3.882	.011
	Content Monitoring	3.335	.014
	Evaluation Monitoring	0.871	.532

4.3. Do different subgroup dynamics in SSRL activities have different impacts on computational thinking performance?

To test the differences in the computational thinking performance of each subgroup, ANOVA statistical tests were carried out. The results (Table 6) showed that computational thinking performance among subgroups was significantly different, $F = 4.495$, $p = .001$. That is, some subgroups had high scores, whereas other subgroups had low scores. For example, subgroup #2 had a mean value of 97.500, whereas subgroup #7 had a mean value of 38.333.

There are two possible reasons for such differences. First, the difference can be accounted for by how students cooperated during SSRL activities. The average frequency of subgroup regulation in SSRL was 13.667 (SD = 4.5019), 16.833 (SD = 1.7224), 10.333 (SD = 4.5898), 9.333 (SD = 5.5737), 12.333 (SD = 4.5461), 10.833 (SD = 5.2964), 4.500 (SD = 4.6797), and 6.250 (SD = 5.5050). After comparing the frequency values of the two subgroups with the best and the worst scores (i.e., subgroups #2 and #7) using *t*-tests, we found (see Table 7) that there was a significant difference, $t = 6.330$, $p = .001$.

Table 6. The computational thinking performance among experimental subgroups

Group	*N*	Mean	*SD*	*F*	*p*
1	6	72.500	8.512	4.495	.001
2	6	97.500			
3	6	59.167			
4	6	61.333			
5	6	65.000			
6	6	67.500			
7	6	38.333			
8	4	40.000			

Table 7. The difference between subgroups #2 and #7 in SSRL activities

Subgroup	Mean	SD	*t*	*Sig.*
2	16.833	1.722	6.330	.001
7	4.500	4.680		

Second, the difference can also be accounted for by how students participated in SSRL activities. A *t*-test was carried out to compare the four SSRL processes between the two subgroups (i.e., #2 and #7), and the results are shown in Table 8. There was a significant difference between subgroups #2 and #7 in the four processes. T and p values for each process were $t = 3.742$, $p = .004$ (planning and goal setting), $t = 4.472$, $p = .001$ (task monitoring), $t = 7.593$, $p = .000$ (content monitoring), and $t = 3.803$, $p = .003$ (evaluation monitoring). The greatest difference between the two subgroups was in content monitoring.

Table 8. The differences between subgroups #2 and #7 in the four SSRL processes

Process	Subgroup	*M*	*SD*	*t*	*Sig.*
Planning and goal setting	2	3.17	0.983	3.742	.004
	7	.83	1.169		
Task monitoring	2	3.00	0.894	4.472	.001
	7	1.00	0.632		
Content monitoring	2	6.00	0.632	7.593	.000
	7	1.33	1.366		
Evaluation monitoring	2	4.33	0.516	3.803	.003
	7	1.33	1.862		

In summary, there are differences in how students cooperate and participate in SSRL activities. Those who cooperated more had better scores. In addition, those who participated in SSRL activities more actively (especially in content monitoring) had better scores.

5. Discussion

5.1. Empirical contributions

The findings of this research demonstrated that the students in the experimental group significantly outperformed the students in the control group on the midtest and posttest. When students were engaged in SSRL activities, their task monitoring and content monitoring processes improved. We also found that the learning gain of the experimental group was significant from the pretest to the midtest.

The results of this study suggest that SSRL activities were beneficial for computational thinking performance. That is, SSRL involves different aspects of learning regulation to ensure that group members remain involved and provide consistent efforts during their learning process. As a result, these students' performance was much better than that of their counterparts. Our findings are supported by the related literature. For example, students in the experimental group were engaged in SSRL activities that advanced their learning outcomes (Hadwin et al., 2011a; Järvelä et al., 2013; Panadero & Järvelä, 2015; Rogat et al., 2011). Throughout the learning process, SSRL includes several processes, such as planning, monitoring and evaluation, in which the students are engaged from the beginning to the completion of their learning process. Through this process, students know what to learn, what homework to complete and how to evaluate their learning progress. Therefore, SSRL improves the pertinence and effectiveness of students' learning through group planning and monitoring as well as final evaluation. As a result, academic performance can be improved with SSRL.

Our results also suggest that regulation in SSRL is more effective during the first half of computational thinking learning. This finding implies that in the early stage of cooperative learning, individuals are not familiar with one another and need to adapt to the cooperative nature of learning. Group cooperative learning regulates individual behavior through engagement and group supervision to complete cooperative tasks and improve academic performance. At the later stage, such group supervision activities become increasingly less obvious. This research result verifies the views of Panadero and Järvelä (2015). These scholars stated that SSRL can contribute to students' learning performance and that coregulation occurs in some periods in groups. This would be the case as groups progress through different phases of their collaboration and do not always socially share regulations of learning (Panadero & Järvelä, 2015). In our research, coregulation occurred in the early stage.

In the process of SSRL, monitoring (i.e., task and content monitoring) plays a major role in student performance. This finding is consistent with the hypothesis that socially shared plans, tasks, and content are important factors for successful collaborative learning (Schoor & Bannert, 2012). The groups with higher learning outcomes tended to participate in socially shared tasks and socially shared content and had more interactive behaviors in completing the tasks. Successful teams put more effort into SSRL planning and task analysis. In the learning process, students need to monitor their learning behavior and learning outcomes, which can be facilitated by group work. The group members regulate learning through discussion, planning, implementation, monitoring and evaluation of learning. This study results prove once again that good regulation of learning is necessary, especially for the completion of learning tasks and the improvement of academic performance.

5.2. Theoretical contributions

There were also differences among subgroups in the SSRL performed. The supervision activities of the more successful groups were relatively stable and occurred more frequently, which was similar to the results of other studies (e.g., Järvelä et al., 2016). The reason for the difference among groups may be accounted for by the following factors. One reason is group norms. In the implementation of SSRL, students had a unified concept and executed it according to the needs of the task, which effectively promoted the generation of students' individual cognitive behavior. Moreover, group norms were beneficial in stimulating group motivation, exerted a positive and significant impact on group cognition, and promoted students' academic performance. In cooperative learning, with the enhancement of the awareness of supervision activities, students' group consciousness gradually formed. The shared learning rules and regulations of group members improved students' communication and cognitive awareness and cultivated students' self-regulation and social regulation. This result is similar to the study of Azevedo (2014). The other possible reason is group cognitive responsibility. In cooperative learning, social regulation ensures or enhances individual responsibility and supports behavior by defining tasks, scheduling and monitoring group processes and building mutual trust (Fransen et al., 2011). Therefore, in teaching, we can improve the instructional effect by arranging group cooperative learning. In particular, in group cooperative learning, team members can be asked to clarify their learning tasks first, plan the learning process, monitor their learning progress, and reflect on and evaluate the whole learning process. In this way, group cooperative learning can become more meaningful, and learning results can be improved. Furthermore, such an arrangement of the learning process will help improve trust among group members and straighten their collaboration.

5.3. Limitations and future directions

Some limitations of this research need to be acknowledged. The experimental data of this study came from the learning content and the learning object, which is specific, and the amount of data collected was limited. The computational thinking scores were calculated as total scores. In future research, computational thinking can be compared in terms of the calculation concept (7 aspects such as data, cycle and condition) and calculation practice (4 aspects such as increment and iteration), and respective conclusions can be drawn. In addition, the records generated by the team can also be used to improve the accuracy of research results by using sequence mining or process mining techniques (Winne & Philip, 2015). The focus on how SSRL appears in time, how it is triggered, and how SSRL may fluctuate in the process of participation needs to be further explored.

References

Aho, A. V. (2012). Computation and computational thinking. *Computer Journal, 55*(7), 832-835.

Angeli, C., Voogt, J., Fluck, A., Webb, M., Cox, M., Malyn-Smith, J., & Zagami, J. (2016). A K-6 computational thinking curriculum framework: Implications for teacher knowledge. *Educational Technology & Society, 19*(3), 47-57.

Azevedo, R. (2014). Issues in dealing with sequential and temporal characteristics of self- and socially-regulated learning. *Metacognition and Learning, 9*(2), 217-228.

Barefootcomputing (2021). *C.A.S. Computational thinking.* Barefootcomputing. http://barefootcas.org.uk/barefoot-primary-computing-resources/concepts/computational-thinking/

Brennan, K., & Resnick, M. (2012, April). *New frameworks for studying and assessing the development of computational thinking*. Paper presented at annual American Educational Research Association meeting, Vancouver, BC, Canada.

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219-243.

Dagienė, V., & Futschek, G. (2008). Bebras international contest on informatics and computer literacy: Criteria for good tasks. In *International conference on informatics in secondary schools-evolution and perspectives* (pp. 19-30). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-69924-8_2

Denner, J., Werner, L., & Campe, S., & Ortiz, E. (2014). Pair programming: Under what conditions is it advantageous for middle school students? *Journal of Research on Technology in Education, 46*(3), 277-296. https://doi.org/10.1080/15391523.2014.888272

Fransen, J., Kirschner, P. A., & Erkens, G. (2011). Mediating team effectiveness in the context of collaborative learning: The Importance of team and task awareness. *Computers in Human Behavior, 27*(3), 1103-1113.

Hadwin, A. F., Järvelä, S., & Miller, M. (2011a). Self-regulation, co-regulation, and socially shared regulation: Exploring perspectives of social in self-regulated learning theory. *Teachers College Record, 113*, 240-264.

Hadwin, A. F., & Oshige, M. (2011b). Self-regulation, coregulation, and socially shared regulation: Exploring perspectives of social in self-regulated learning theory. *Teachers College Record, 113*, 240-264.

Hadwin, A. F., Järvelä, S., & Miller, M. (2017). Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (pp. 83–106). Routledge/Taylor & Francis Group.

Hu, H., & Driscoll, M. P. (2013). Self-Regulation in e-learning environments: A Remedy for community college? *Educational Technology & Society, 16*(4), 171–184.

Hwang, W. Y., Hsu, J. L., Shadiev, R., Chang, C. L., & Huang, Y. M. (2015). Employing self-assessment, journaling, and peer sharing to enhance learning from an online course. *Journal of Computing in Higher Education, 27*(2), 114–133.

Hwang, W. Y., Shadiev, R., Wang, C. Y., & Huang, Z. H. (2012). A Pilot study of cooperative programming learning behavior and its relationship with students' learning performance. *Computers & Education, 58*(4), 1267-1281.

Ge, M. (2014). 基于计算思维的协作学习模式在中学信息技术课程中的实践与研究 [A Study about the practice of collaborative learning mode based on computational thinking in middle school information technology curriculum]. https://cdmd.cnki.com.cn/Article/CDMD-10718-1014402627.html

Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. *Educational Psychologist, 48*(1), 25-39.

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.

International Society for Technology in Education (2021). *Computational Thinking Competencies*. https://www.iste.org/standards/computational-thinking

Lee, A. (2014). *Socially shared regulation in computer-supported collaborative learning.* (Publication No. 3641808) [Doctoral dissertation, Rutgers University-Graduate School-New Brunswick]. ProQuest Dissertations and These Global.

Malmberg, J., Järvelä, S., & Järvenoja, H. (2017). Capturing temporal and sequential patterns of self-, co-, and socially shared regulation in the context of collaborative learning. *Contemporary Educational Psychology, 49*, 160-174.

Maria, S., & Tsiatsos, T. (2017). Case study: Integrating computational thinking into the introductory course of computer science via the use of the programming language Python. In *Interactive Mobile Communication, Technologies and Learning* (pp. 531-541). https://doi.org/10.1007/978-3-319-75175-7_52

Mcdowell, C., Werner, L., & Bullock, H. E., & Fernald, J. (2003). The Impact of pair programming on student performance, perception and persistence. In *International Conference on Software Engineering* (pp. 602-607). https://doi.org/10.1109/ICSE.2003.1201243

Miller, M., Järvelä, S., & Hadwin, A. (2017). Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of Self-Regulation of Learning and Performance* (pp. 83–106). New York, NY: Routledge.

Panadero, E., & Järvelä, S. (2015). Socially shared regulation of learning: A Review. *European Psychologist, 20*(3), 190-203.

Rogat, T. K., & Linnenbrink-Garcia, L. (2011). Socially shared regulation in collaborative groups: An Analysis of the interplay between quality of social regulation and group processes. *Cognition and Instruction, 29*(4), 375-415.

Román-González, M., Moreno-León, J., & Robles, G. (2019). Combining assessment tools for a comprehensive evaluation of computational thinking interventions. In *Computational Thinking Education* (pp. 79-98). https://doi.org/10.1007/978-981-13-6528-7_6

Schoor, C., & Bannert, M. (2012). Exploring regulatory processes during a computer-supported collaborative learning task using process mining. *Computers in Human Behavior, 28*(4), 1321-1331.

Shadiev, R., Hwang, W. Y., Yeh, S. C., Yang, S. J. H., Wang, J. L., Lin, H., & Hsu, G. L. (2014). Effects of unidirectional vs. reciprocal teaching strategies on web-based computer programming learning. *Journal of Educational Computing Research, 50*(1), 67-95.

Turchi, T., Fogli, D., & Malizia, A. (2019). Fostering computational thinking through collaborative game-based learning. *Multimedia Tools and Applications, 78*(10), 13649-13673.

Vauras, M., Iiskala, T, & Kajamies, A., Kinnunen, R., & Lehtinen, E. (2003). Shared-regulation and motivation of collaborating peers: A Case analysis. *Psychologia, 46*(1), 19-37. https://doi.org/10.2117/psysoc.2003.19

Volet, S., Vauras, M., Khosa, D., & Iiskala, T. (2013). Metacognitive regulation in collaborative learning: conceptual developments and methodological contextualizations. In *Interpersonal Regulation of Learning and Motivation* (pp. 81-115). Routledge.

Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33-35.

Winne, & Philip H. (2015). What is the state of the art in self-, co- and socially shared regulation in CSCL? *Computers in Human Behavior, 52*, 628-631.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated engagement in learning. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in Educational Theory and Practice* (pp. 277–304). Hillsdale, NJ: Erlbaum

Winner, P. H. (2011). A Cognitive and metacognitive analysis of self-regulated learning. In B. J. Zimmermanand & D.H. Shunk (Eds.), *Handbook of Self-Regulation of Learning and Performance* (pp. 15-32). New York, NY: Routledge.

Yağcı, M. (2019). A Valid and reliable tool for examining computational thinking skills. *Education and Information Technologies, 24*(1), 929-951.

Zheng, L., Li, X., & Huang, R. (2017). The Effect of socially shared regulation approach on learning performance in computer-supported collaborative learning. *Educational Technology & Society, 20*(4), 35–46.

Zheng, J., & Xing, W. & Zhu, G. (2019). Examining sequential patterns of self- and socially shared regulation of STEM learning in a CSCL environment. *Computers & Education, 136*, 34-48.

Appendix

A few examples of test questions

A. Calculation concept

A1. Sequences

To have time for dinner, Sary (S) needs to communicate with five classmates: Alice (A), Bean (B), Cary (C), David (D) and Emil (E). S can communicate with E immediately. However, there are a few things that need to be clear about communicating with her classmates:
1. Before she talks to D, she must communicate with A.
2. Before she talks to B, she must communicate with E.
3. Before she talks to C, she must communicate with B and D.
4. Before she talks to A, she must communicate with B and E.

Question: If, according to the above requirements, Sary hopes to chat with all the above friends, what order should she follow?
A: E, B, A, D, C
B: B, E, A, D, C
C: E, A, B, D, C
D: E, A, B, D, C

A2. Data and operators

Grandma Fox does not know how to use a computer. However, she had to set a password for her mailbox to keep it safe, and Grandma Fox must follow the following requirements to set a password:
1. A minimum of 2 letters must be capitalized.
2. There must be more English letters than Arabic numerals.
3. Minimum 3 special characters (neither English nor numeric).

Question:
Which of the following passwords conforms to the above rules?
A: PearL@mb2953?
B: ##RedM3rgan-2688
C: R5#X&v73r68!?
D: *h9n3ytR33*§!

A3. Conditionals

The Styx operating system has a feature in which a poisoned Styx operating system computer will return an incorrect answer to any question received from the Internet. If it is asked, "Are you infected with the virus?" It will answer "no." An uninfected computer always answered correctly with "no" when asked, "Have you been infected?" Styx's information engineers tested Styx servers and laptops over the Internet.
Question:
In which of the following sentences is a message returned only by a poisoned server?
A: I am a poisoned server.
B: I am not a poisoned server.
C: I am a poisoned laptop.
D: I am not a poisoned laptop.

B. Calculation practice

B1. Reusing and recreating

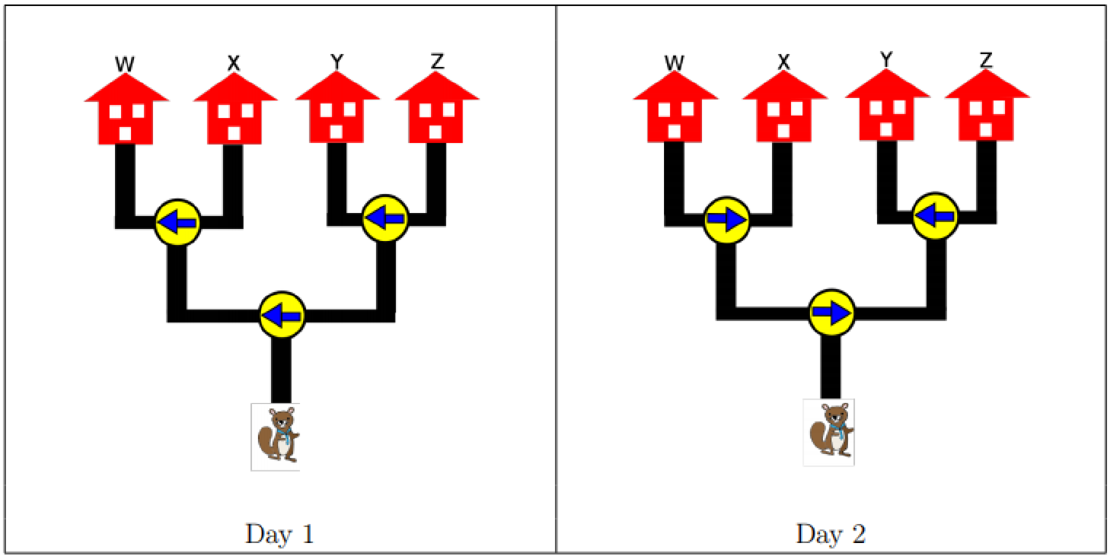
You have a beautiful paper airplane (one of those things we often fold), and then you need to transform or recreate it into a new shape (airplane or other object). What would you do? (multiple choice)
A: Take the old one apart, fold it and try a new idea over and over again.
B: Take the folded plane and make a slight change.
C: Follow the feeling and try new folding methods again and again.
D: Do not want to change the original, feel the original plane has been very good.

B2. Abstracting and modularizing

Class task: Let's use information technology to create an "explore the moon" handwritten newspaper. In what way do you think we can best accomplish this task? (multiple choice)

A: Break down the handwritten newspaper into several columns and finish it one by one.
B: Use search engines to search for good templates and then modify them.
C: Design and make based on the production experience in class.
D: Ask teachers for help and let them guide me in completing it.

B3. Incremental and iterative
Mr. Beaver has 4 friends living in different villages, and he plans to visit one of these friends every afternoon. Mr. Beaver will follow the direction of the arrow on signs at each intersection. Initially, all arrows point to the left road. When passing an intersection, Mr. Beaver switches the arrow to the opposite direction. For example, on Day 1, Mr. Beaver takes the road on the left at the first intersection, takes the left road on the second intersection, and reaches Village W. On Day 2, Mr. Beaver turns right at the first intersection, then left at the second intersection, and arrives at Village Y.



Question:
Which village will Mr. Beaver visit on day 30?
A: Village W
B: Village X
C: Village Y
D: Village Z

Chang, H.-T., Lin, C.-Y., Wang, L.-W., & Tseng, F.-C. (2022). How Students can Effectively Choose the Right Courses: Building a Recommendation System to Assist Students in Choosing Courses Adaptively. *Educational Technology & Society*, *25* (1), 61-74.

How Students can Effectively Choose the Right Courses: Building a Recommendation System to Assist Students in Choosing Courses Adaptively

Hui-Tzu Chang[1], Chia-Yu Lin[2*], Li-Chun Wang[3] and Fang-Ching Tseng[4]
[1]Center for Institutional Research and Data Analytics, National Yang Ming Chiao Tung University, Hsinchu // [2]Department of Computer Science and Engineering, Yuan Ze University, Taoyuan // [3]Department of Electrical and Computer Engineering, National Yang Ming Chiao Tung University, Hsinchu // [4]Electrical Engineering and Computer Science Undergraduate Honors Program, National Yang Ming Chiao Tung University, Hsinchu // simple@nycu.edu.tw // sallylin0121@gmail.com // wang@nycu.edu.tw // claire.eecs07@nctu.edu.tw
[*]Corresponding author

(Submitted March 3, 2021; Revised June 21, 2021; Accepted August 1, 2021)

ABSTRACT: In this study, we built a personalized hybrid course recommendation system (PHCRS) that considers students' interests, abilities and career development. To meet students' individual needs, we adopted the five most widely used algorithms, including content-based filtering, popularity-based methods, item-based collaborative filtering, user-based collaborative filtering, and score-based methods, to build a PHCRS. First, we collected course syllabi and labeled each course (e.g., knowledge/skills taught, basic/advanced level). Next, we used course labels and students' past course selections and grades to train five recommendation models. To evaluate the accuracy of the system, we performed experiments with students in the Department of Electrical and Computer Engineering, which provides 1794 courses for 925 students and utilizes the receiver operating characteristic curve (ROC) and normalized discounted cumulative gain (NDCG) as metrics. The results showed that our proposed system can achieve accuracies of 80% for ROC and 90% for NDCG. We invited 46 participants to test our system and complete a questionnaire. Overall, 60 to 70% of participants were interested in the recommended courses, while the course recommendation lists produced by content-based filtering were in line with 67.40% of students' actual course preferences. This study also found that students were more interested in courses at the top of the recommendation lists, and more students were autonomously motivated than held extrinsic informational motivation across the five recommendation methods. These findings highlighted that the proposed course recommendation system can help students choose the courses that interest them most.

Keywords: Course recommendation, Course selection, Learning aids, Personalized learning

1. Introduction

Studying at a university involves taking a wide variety of optional courses, especially for students in larger departments, and students have to carefully consider which of the numerous optional courses would be best for them to take. Course options are important for fulfilling degree requirements and determining future careers (Farzan & Brusilovsky, 2006; Kurniadi et al., 2019). Given the large number of optional courses, students may need to dedicate a great deal of time to researching information for each course to select the best options for their situation. Since students do not always have enough information, it can be challenging for them to make the right decision (Chang et al., 2020; Wang et al., 2020); students are often influenced by the opinions of other students. Under these conditions, it is important to collect student and course information and then perform further analysis to determine which courses might meet each student's personal needs. One solution would be through a course recommendation system that helps students make a good decision (Iatrellis et al., 2017; Sawarkar et al., 2018).

Course recommendation systems use different techniques to collect students' past educational data and then automatically provide course match predictions and recommendations by analyzing the data (Aguilar et al., 2017; Romero & Ventura, 2013). The collaborative filtering method (Chang et al., 2020; Wang et al., 2020), demographic-based filtering method (Dwivedi & Roshni VS, 2017; Zhang et al., 2015), content-based filtering method (Apaza et al., 2014; Esteban et al., 2020), and knowledge-based filtering method (Aher & Lobo, 2013; Kurniadi et al., 2019) are common methods used in the existing recommendation systems, although no existing course recommendation system uses more than one of these methods. Since each student has different motivations and different needs from optional courses, different recommendation methods should be combined into a single recommendation system. In addition, all recommendation methods have positive and negative aspects. To mitigate any disadvantages, many systems choose to use hybrid recommendation methods (Çano & Morisio, 2017; Zhang et al., 2015). Therefore, in this paper, we propose a personalized hybrid course recommendation system (PHCRS) that integrates five recommendation methods for formal offline courses to consider the different learning needs of students.

The goal of the PHCRS is to provide information based on students' preferences; however, current systems mostly focus on how to improve students' grades (Esteban et al., 2020) and help students achieve their long-term career goals (Farzan & Brusilovsky, 2006). These results-oriented PHCRSs do not consider factors that affect students' course-selection process. If a system aims to provide personal recommendations, it is important to fully understand the factors affecting the reasoning behind optional course selection (Han et al., 2016) and then provide many recommendation methods for the students to choose from. Thus, the main purpose of this study is to construct a PHCRS that takes students' interests, abilities, and careers into consideration and then provide a course recommendation list based on their preferred fields to satisfy the need to select optional courses. Our study also evaluated the accuracy of the recommendation model and then empirically assessed whether students were interested in the course recommendation list provided by numerous recommendation methods and the factors affecting that interest. We will eventually expand and modify the system functionality to fit the needs of the students.

2. Literature review

2.1. Students choosing courses

Course selection is regarded as an important aspect in student's experiences of university. Students need to make a series of course selection decisions before the semester starts and these decisions have a decisive effect on their future life, education, and employment opportunities (Babad & Tayeb, 2003). The course decision-making process is affected by many factors, and there is usually no perfect combination of courses, as some factors may lead to conflicting demands (Lang, 2010). The resulting issues may interfere with students' judgment, as students commonly use their instincts or information provided by others to choose their courses (Babad et al., 2004), and these decisions affect their learning experiences. Choosing a major is the most important decision when entering university (Begs et al., 2008). Perera and Pratheesh (2018) found that major selection is affected mostly by career factors, and academic quality, personality, and ability also affect this decision. This results in students not always choosing majors they are interested in. Zare-ee and Shekarey (2010) also found that family, social, and personal factors, such as parental educational level, household income, media use, GPA, and personal interests, may force students to change their minds.

After students have decided their majors, they choose the courses they will take each semester in accordance with school regulations. Babad et al. (2004) proposed a theory of students' course selection as a decision-making process using the dimensions of learning values, learning styles, and course difficulty. Babad et al. (2004) found that the importance of academic intelligence and teachers' lecture style are key components affecting students' course selection. Babad (2001) also found that students' first course selection decisions are based mostly on the course's content, lecture quality and potential value for future careers. Conversely, the last course selection decisions are usually based on course difficulty (the easier the better), and a comprehensive course selection strategy may reduce risk in course selection. On the other hand, determining course selection motivations is a more complicated process and includes both autonomous motivation and extrinsic informational motivation (Lee & Sun, 2010). The former is a spontaneous behavior generated by the self-motivated interest, curiousness or career planning of an individual, and the latter is a behavior influenced by the external environment, such as the desire for certain grades, rewards or ratings. Students who select courses based on their autonomous motivation tend to be more devoted to studying than those who select courses based on extrinsic informational motivation (Lee & Sun, 2010). Thus, it is clear that when students select courses, major, course importance and difficulty, and course selection motivations are their main concerns.

2.2. Recommendation system

Recommendation systems originated in e-commerce recommend products based on user preferences (Burke, 2002). These have become a fundamental part of e-commerce, requiring massive information collection, analysis, and prediction. Recommendation systems help users choose the most appropriate products on the basis of their demands and preferences (Resnick & Varian, 1997; Xiao et al., 2018). Notable examples are the systems used by Netflix, Google News, and Amazon (Han et al., 2016). These enterprises use recommendation systems to discover the latent relationships between their items and users and to exploit potential customer demands. They have successfully connected information with sales and helped customers find items they are interested in while also raising the total revenue of the enterprise.

Several recommendation techniques based on different user needs are employed in recommendation systems. (1) Collaborative filtering: This analyzes the similarity between users and items to predict what content users may be interested in (using population characteristics or search history) and recommend it to the user (Burke, 2002; Salehudin et al., 2019). (2) Demographics-based recommendation: This utilizes users' basic information to identify user similarities and then recommends items that have been recommended to users with similar characteristics, such as age and gender (Aguilar et al., 2017; Burke, 2002). (3) Content-based filtering: This system matches item characteristics and user attributes and then searches for items similar to those users expressed previous interest in. This is known as item-to-item similarity (Schafer et al., 1999). (4) Utility-based filtering: These recommendations are based on the match between the demands of a user and available items (Burke, 2002). (5) Knowledge-based filtering: This is an inferencing technique that is based not on user demands or preferences but on differences in functional knowledge. The development of this system requires catalog knowledge, functional knowledge and user knowledge (Aguilar et al., 2017; Kurniadi et al., 2019).

2.3. Recommendation systems in education

Course options in universities are highly related to career development. In the institutional education process, college is an important transition period for students. Seventy-five percent of college students have not decided what career they want to pursue in the future or even what they want to gain expertise in. Fifty to 75% of students change their major at least once during their time in university (Cuseo, 2003; Gordon, 2007). The main challenge in developing a suitable recommendation system for selected courses is that it is hard to integrate data from different sources. It is also difficult to find effective, useful and precise information online about students' study plans (Obeid et al., 2018). Many students select optional courses without seeking help or advice from outside educational services, which may lead to their skills, interests, and career development plans not aligning with the courses they select, subsequently leading to a decreasing retention rate (Kongsakun et al., 2010). Archer and Cooper (1998) pointed out that university-provided advisory services are important for student success. These services help students determine a study plan, provide career guidance, assist with interpersonal relationship management, and provide an understanding of the physical and mental status of students (Urata & Takano, 2003). Most higher education institutes lack sufficient human resources and talent (Kongsakun et al., 2010). Some schools have asked staff to take on more responsibilities, but they usually do not have enough time to provide complete advisory services, nor do they have enough tools to help needy students (Salehudin et al, 2019). To solve this problem, many schools have tried to utilize recommendation systems to provide support for students' decision making (Aher & Lobo, 2013; Bendakir & Aïmeur, 2006; Romero & Ventura, 2013).

As technology progresses, learners will contribute more to data collection through learning platforms by browsing courses, interacting with the interface, and requesting records. The large amount of data collected contains information on the implicit intentions, interests, and educational performance of students. If the recommendation system can utilize these data to guide students toward suitable learning opportunities, it can help meet students' learning needs (Aguilar et al., 2017). In recent years, course recommendation systems have been developed. A course recommendation system analyzes the selected data and then combines it with past student data to automatically predict preferences and provide recommendations through education data analysis (Aguilar et al., 2017).Using recommendation systems to guide students in their educational decisions has a significantly positive effect (Kurniadi et al., 2019).

Xu (2016) proposed a course sequence recommendation system to reduce students' time to graduation and maximize their performance. This system analyzed the prerequisite dependency among courses to adaptively recommend online learning course sequences to students. Hou et al. (2018) designed a contextual recommendation system to solve heterogeneity issues in large-scale user groups and sequencing issues regarding online learning courses. In the paper (Mondal et al., 2020), the authors combined K-means clustering and collaborative filtering techniques to propose an online course recommendation system based on grades. These studies chose certain online learning university courses such as massive open online courses (MOOCs) for which to implement recommendation systems since online learning is much easier to collect data about than formal higher education courses. However, formal offline courses are more important than online learning courses to students. The variety of formal offline courses offered by universities and the range of skills they teach are wider and more complete than those of online courses. Thus, students have a greater need for recommendation systems when choosing formal offline courses. Yao (2017) developed an intelligent personalized context-aware recommendation (PCAR) learning system to recommend suitable learning materials from various learning environments. Huang et al. (2019) designed a cross-user-domain collaborative filtering algorithm to recommend optional courses for college students by accurately predicting the interest they would have in optional courses. Pardos et al. (2020) built course2vec models based on course catalog descriptions and enrollment histories to prepare an appropriate recommendation system for the university context. Ultimately, all of these works implemented recommendation systems for formal offline courses.

The above works show that common methods for developing the recommendation system are as follows:

- Collaborative filtering. This includes both item-based and user-based filtering methods. Item-based filtering uses students' grades in other subjects to recommend courses (Chang et al., 2020; Dwivedi & Roshni VS, 2017; Wang et al., 2020). User-based filtering matches a student's course selection route with alumni who shared a similar route and recommends the course list of the alumni to the student (Bendakir & Aimeur, 2006; Perugini et al., 2004; Zhang et al., 2015).
- Demographic-based filtering. This method draws upon population characteristics to classify recommendation demands of different groups. It recommends courses that a group may be interested in on the basis of the age, gender or intended or previous profession of the students. This method is mostly used in MOOC open courses (Dwivedi & Roshni VS, 2017; Zhang et al., 2015) and lifelong learning courses (Han et al., 2016; Tuckman, 1999).
- Content-based filtering. This method is based on characteristics listed in course syllabi, such as the subject field and lecture content. The system is able to provide a course list to a student that is similar to his/her past course list (Apaza et al., 2014; Esteban et al., 2020; Herlocker et al., 2000).
- Knowledge-based filtering. This may use students' past grades to determine courses for which they might receive similar results. Alternatively, it may analyze the students' overall GPA and then use the recommendation results to predict students' future grades or likelihood of graduation. Based on the results, the system then provides a list of the most suitable courses to students (Aher & Lobo, 2013; Kurniadi et al., 2019).

However, each of the currently existing formal offline course recommendation systems uses only one of these recommendation methods. A robust recommendation system should combine different recommendation methods to provide diverse suggestions since each student has different motivations and preferences when choosing courses. Additionally, all recommendation methods have both positive and negative aspects. To mitigate any disadvantages, many systems use hybrid recommendation methods (Çano & Morisio, 2017; Zhang et al., 2015). Of all methods available, the collaborative filtering & content-based filtering hybrid recommendation method is the most common (Esteban et al., 2020). It overcomes the limitations of both collaborative filtering and content-based filtering methods and increases predictability while also decreasing the degree of sparsity and loss of information. Therefore, we propose a PHCRS including five recommendation methods for formal offline courses to consider the different learning needs of students.

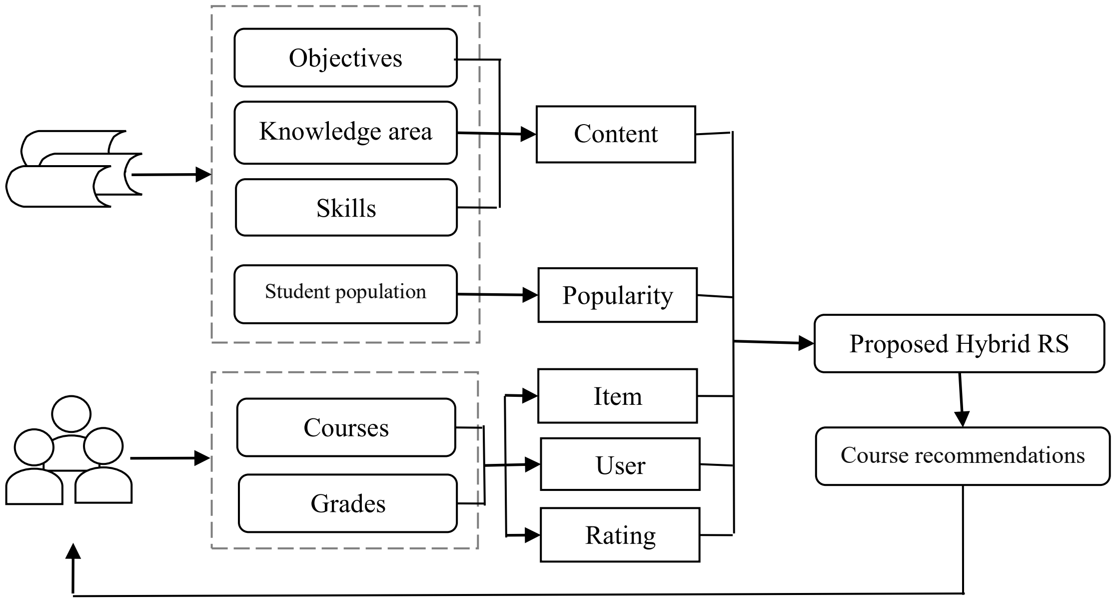
Based on the literature mentioned above, students' final course decisions are affected by their major and motivation as well as school requirements. Esteban et al. (2020) suggest that students' personal characteristics, such as major, learning goals, and desires, should be taken into consideration when developing a course recommendation system in order to provide tailored course recommendations to students. This research proposes four hypotheses to verify the effectiveness of the PHCRS:

- Hypothesis 1: Students' degree of interest in the courses recommended by the five recommendation methods will differ among the undeclared field and three optional fields.
- Hypothesis 2: Students' degree of interest in the courses recommended will differ according to the order of the recommended courses.
- Hypothesis 3: The degree of interest in the courses recommended to a student will be affected by the student's internal and external motivations for taking a course.
- Hypothesis 4: The degree of interest a student has in the recommended courses will vary with the recommendation methods used and their degree of suitability for the student.

3. Development of a system for recommending adaptive courses

This research proposes a PHCRS, as shown in Figure 1. We use courses and student data from the Big Data Research Center in National Yang Ming Chiao Tung University (NYCU) to train the recommendation system. These data contain information on 386 different courses from the Department of Electrical and Computer Engineering, and a total of 2985 courses were provided from the fall 2011 semester to the fall 2020 semester. For student information, a total of 1824 students from the Department of Electrical and Computer Engineering who were enrolled between 2011 and 2020 were selected. To prepare the training data, the researchers collected the course outlines and interviewed the teachers via telephone. The two researchers discussed and agreed upon the labeling rules and then compared the similarities and differences in the labeling results after making the labels. In cases of disagreement, the scorers discussed the issue until a consensus was reached. The interrater reliability fell between .7 and .8. The attributes of each course was labeled as follows. (1) Course objectives: This label indicates what the course mainly teaches students, such as signal processing or communication systems. There are a total of 44 possible labels. (2) Knowledge areas: This label is based on the theories, methods or empirical theories from the field of electrical engineering that are taught to students, such as information and communication, system-on-chip, and 13 other areas. (3) Skills: This label is based on the relevant technologies, resources or tools used in each course, such as Python or MOSFET. There are a total of 203 possible labels. After the data preparation, five recommendation methods were implemented in HPCRS for students with different learning needs as follows:

Figure 1. PHCRS



3.1. Recommendation model construction

- Content-based Filtering: Content-based filtering recommends similar courses based on the characteristics of students' past courses (Esteban et al., 2020). In the first step, the feature vectors of the courses is extracted. The course feature vector indicates which domains the courses belong to and which objectives the courses contain. To calculate the feature vectors of *student x* for *course i,* the feature vector of *course i* is multiplied by the score of the *student x* on *course i*. We add up all the feature vectors of *student x* on each course and define this value as the feature vector of student *x*. To recommend *course j* to *student x*, we use the feature vector of student *x* and the feature vector of *course j* to calculate cosine value ($\cos\theta = \frac{\vec{i}\cdot\vec{j}}{||\vec{i}||\cdot||\vec{j}||}$) as the similarity. If the similarity is close to 1, *student x* is more likely to like *course j*.
- Popularity-based Method: Popularity-based method counts the number of students in each course, and recommend the course with the largest number of students. (Burke, 2002).
- Item-based Collaborative Filtering: Item-based collaborative filtering calculates the similarity score between courses and recommend similar courses (Sarwar et al., 2001). We find the students who have taken these two courses and calculate the difference of their scores in the two courses. The smaller the difference, the higher the similarity. The similarity is represented as $w_{i,j}$ and is shown in (1), where A are the set of students who have taken *course i* and *course j*. Assuming *student x* has taken *course i*, if PHCRS want to recommend *course k* to *student x*, the predicted score is calculated by formula (2). The numerator is equal to the product of $w_{i,k}$ and the student's grade in *course i*. The denominator is the summation of the similarity between *course i* and *course k*.

$$\text{Similarity between } \textit{course i} \text{ and } \textit{course j } (w_{i,j}) = \frac{1}{1+\sqrt{\sum_{A\in M(i)\cap M(j)}(grade(A,i)-grade(A,j))^2}} \quad (1)$$

$$\text{The prediction score of } \textit{course } k \text{ for } \textit{student } x = \frac{\sum_{w_{i,k}>0} grade(x,i) * w_{i,k}}{\sum_{w_{i,k}>0} w_{i,k}} \quad (2)$$

- User-based Collaborative Filtering: User-based collaborative filtering utilizes students' past course data to calculate the similarity between students and recommend courses taken by similar students (Han et al., 2016). To calculate the similarity between two students, we have to find out the courses the students have both taken. We utilize the scores of two students in the courses to calculate the similarity. The similarity of *student x* and *student y* is represented as weighted value ($w_{x,y}$) and is shown in (3), where *N(x)* are the courses that *student x* has taken, and *N(y)* are the courses that *student y* has taken. If the scores are closer, the similarity of two students is higher. If PHCRS want to recommend *course k* to *student y*, the similarity of *student x* and *student y* is multiplied by the scores of *student x* on *course k*. The average of weighted value is the predicted score, as shown in (4).

$$\text{Similarity of } \textit{student } x \text{ and } \textit{student } y \ (w_{x,y}) = \frac{1}{1+\sqrt{\sum_{i\in N(x)\cap N(y)}(grade(x,i)-grade(y,i))^2}} \quad (3)$$

$$\text{The predicted score for } \textit{student } y \text{ on } \textit{course } k = \frac{\sum_{w_{x,y}>0.2} grade(y,k) * w_{x,y}}{\sum_{w_{x,y}>0.2} w_{x,y}} \quad (4)$$

- Score-based Method: Score-based method calculates the total average score of the class for each course and recommend the course with the highest average score (Sawarkar et al., 2018).

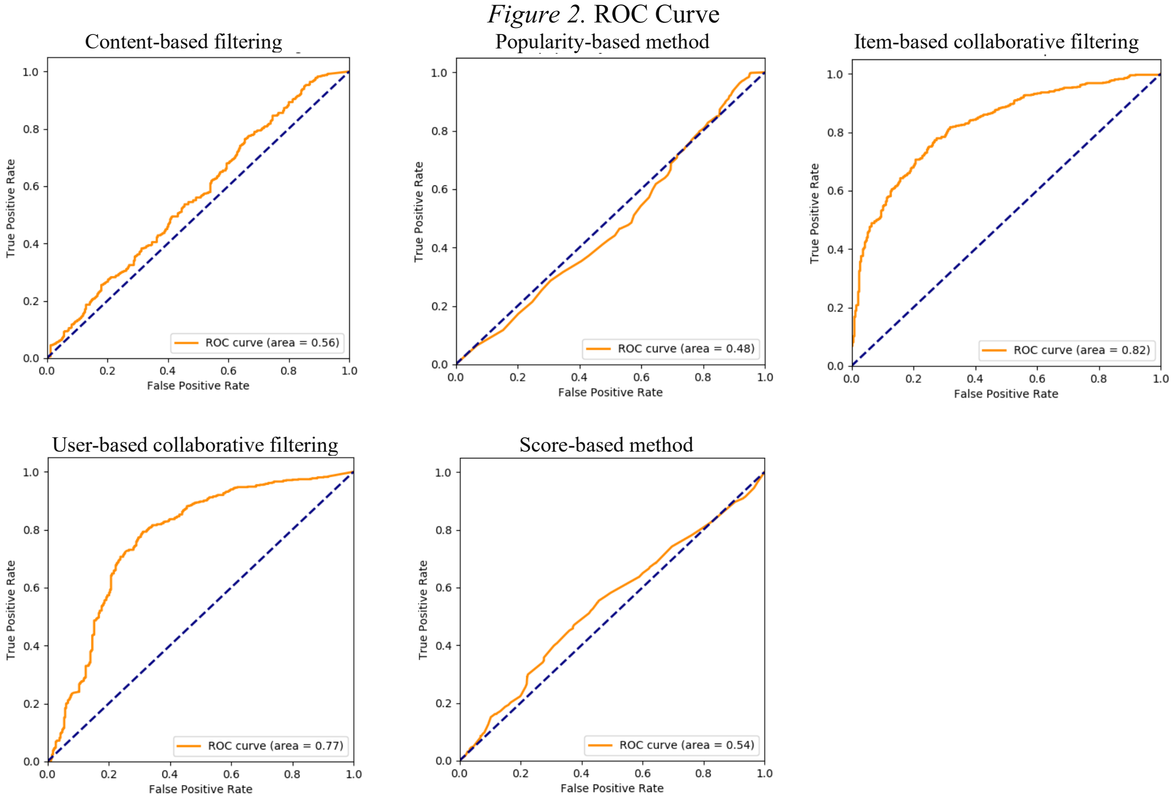
3.2. Evaluation of the recommendation results

This study uses the receiver operating characteristic curve (ROC) and NDCG to evaluate the recommendation results.

3.2.1. ROC

ROC is a coordinate diagram analysis tool used to select the best signal detection model and is also often being used for evaluation of recommendation systems (Zweig & Campbell, 1993). We use the grades of students as the evaluation indicator. For each course, we calculate the average score of students who have taken the course. If the grade of a student is higher than the average score, we say that the course is suitable for the student and call it "true value." For each student and each course, the recommendation system predicts a score for the student. For all test data, we will get many predicted scores. We take every predicted score as the threshold to draw ROC. If the predicted score of the course for the student is higher than the threshold, it would be judged as positive. Otherwise, it is negative. Therefore, we can compare the ground truth and the predicated result to get true positive (TP), false positive (FP), true negative (TN), and false negative (FN). The true positive rate (TPR) is TP/(TP+FN). The false positive rate (FPR) is FP/(FP+TN). The ROC curve takes the false positives rate (FPR) as the x-coordinate and true positive rate (TPR) as the y-coordinate. If the area under the ROC curve is above 0.9, the system is highly accurate; whereas if it is between 0.7 and 0.9, this means medium accuracy. If it is between 0.5 and 0.7, this will mean low accuracy and finally poor accuracy can be identified with results below 0.5. This study uses ROC to evaluate the five recommended methods with the best coefficients as item-based collaborative filtering = .82, followed by user-nased collaborative filtering = .77, content-based filtering = .56, score-based method = .54, and popularity-based method = .48 (see Figure 2).

Figure 2. ROC Curve



3.2.2. NDCG

This study uses NDCG to evaluate the five recommendation methods. For k courses, we sort the courses by the recommendation scores and calculate discounted cumulative gain (DCG). The DCG is shown in (5), where k represents the number of courses the system is recommended and rel_i is gain for each recommendation course. In the evaluation, when the recommendation course overlaps the real record, we set the gain rel_i to be 1; otherwise be 0. The ideal course order based on the predicted score is used to calculate ideal discounted cumulative gain (IDCG), as shown in (6). We can use DCG and IDCG to calculate NDCG, as shown in (7).

$$DCG_k = \sum_{i=1}^{k} \frac{2^{rel_i}-1}{log_2(i+1)} \quad (5)$$

$$IDCG_k = \sum_{i=1}^{|rel_k|} \frac{2^{rel_i}-1}{log_2(i+1)} \quad (6)$$

$$NDCG_k = \frac{DCG_k}{IDCG_k} \quad (7)$$

The NDCG of five recommendation methods are: user-based collaborative filtering = .96, popularity-based method = .94, score-based method = .94, content-based filtering = .89, and item-based collaborative filtering = .89.

3.3. Building a recommendation system website

Our course recommendation website was built using WordPress (WordPress.com, 2021) and is hosted on Xammp (Apache Friends. 2021). The website is embedded in the NYCU portal so that both student and course information data can be updated before the course selection period in each semester. To prevent data overload and to enhance the performance of the website, we imported the data into the website database after it was computed and simplified. The two main features of our website include personal learning analysis and course recommendation services (see Figure 3). The personal learning analysis helps students understand their autonomy index and conformity index in course selection, while the course recommendation feature allows them to search for suitable courses by entering their preferences into the recommendation system. The course recommendation website then indicates the suitability of the courses for the student as well as the course name, lecture time and lecturer name.

Figure 3. Screenshot of the course recommendation service system

4. Research design

This study used a survey method to verify the accuracy of the recommendation system. The survey used nonprobability sampling to invite undergraduates from the Department of Electrical and Computer Engineering, NYCU, who volunteered as participants. As the freshmen's course selection and grade data were not yet completed, they were excluded to avoid interference in the research results. A total of 46 students were selected (15 sophomores, 13 juniors, and 18 seniors; 35 males and 11 females). In this research, recruitment posters were sent out by the Department of Electrical and Computer Engineering. After the students signed up, the researchers explained the research process and parameters via phone or mail. To collect data, students were required to log in to the course recommendation system. After reading the description of each recommendation method, students were asked to evaluate whether the courses recommended by each method were of interest and to provide their reasoning. Students could see the overall results for all knowledge fields, and they could choose up to three fields that most interested them. Finally, they were asked to fill in their personal information and offer suggestions for the system.

This study used a recommendation effect scale defined by our research group. When students browsed the course recommendation list, they were asked to evaluate whether each course was of interest to them and the reasons for their answer. For example, when students answered "yes," they would select from a reasons aligned with "autonomous motivation," which comes from careful consideration and self-determination (Lee & Sun, 2010) and includes reasons such as the practicality of the course content, individual learning plans and personal interests, Or reasons aligned with passive "external information motivation" (Lee & Sun, 2010), which included reasons such as making up for missed credits, the course being easy to pass, and seeing good reviews of the teacher. Conversely, if the student answers "No," he or she must also select the reasons for this choice. The options for "autonomous motivation" include the course not being part of their plan, understanding the course content and having no interest in the course. Options for "external information motivation" include the course being too hard, seeing bad reviews of the teacher, and having peers who did not choose the course. The students' overall choice is indicated by "Yes" or "No," and the students can select multiple reasons.

5. Data analysis and results

5.1. An analysis of the differences among the degree of interest in the courses recommended by the five recommendation methods in the undeclared and three optional fields

Repeated-measures ANOVA is used in this section. The data followed a normal distribution (skewness between -1.01 and .49; kurtosis between -1.22 and 1.90). Table 1 shows that the score-based method produced significant differences ($p < .05$), with the first field ($M = 73.27$), second field ($M = 63.09$), and third field ($M = 64.64$) being higher than the undeclared field (Non Field, $M = 55.72$). The results indicated that students were more interested in their optional field course than with the undeclared field courses recommended by the ratings-based method.

Table 1. A differences analysis between the degrees of interest in the courses recommended among the undeclared field and three optional fields

Recommendation methods	Non field		First field		Second field		Third field		F	Multiple comparison
	M	SD	M	SD	M	SD	M	SD		
Content-based Filtering	68.12	24.82	79.88	20.92	75.06	17.38	63.54	26.77	2.62	-
Popularity-based Method	65.51	29.42	71.91	22.53	68.46	30.02	65.87	33.46	.15	-
Item-based Collaborative Filtering	64.20	25.97	67.04	25.29	64.88	33.12	62.13	28.38	.14	-
User-based Collaborative Filtering	61.16	28.71	77.78	25.40	67.16	24.58	65.00	21.22	2.98	-
Score-based Method	55.72	30.85	73.27	27.22	63.09	25.84	64.64	27.70	3.78^{*}	First>Non Second>Non Third>Non

Note. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

5.2. An analysis of the difference among the students' degree of interest in the courses recommended according to the order of the recommendations

Repeated-measures ANOVA is used again in this section. The data follow a normal distribution (skewness between .34 and 1.81; kurtosis between -.77 and 3.46).

Table 2. A differences analysis between the students' degrees of interest in the courses recommended in the course recommendation order

Recommendation methods	First course		Second course		Third course		Fourth course		Fifth course		F	Multiple comparison
	M	SD	M	SD	M	SD	M	SD	M	SD		
Content-based Filtering	87.13	4.68	83.71	14.86	78.72	9.11	67.45	14.86	43.27	13.96	10.25^{***}	First＞Fifth Second＞Fifth Third＞Fifth
Popularity-based Method	83.25	5.44	72.07	3.48	76.85	9.55	68.67	9.93	65.42	10.20	2.75	-
Item-based Collaborative Filtering	79.45	8.72	71.09	7.48	71.10	7.05	68.45	8.51	56.74	4.55	5.94^{**}	First＞Fifth Second＞Fifth Third＞Fifth
User-based Collaborative Filtering	69.51	8.30	72.74	5.54	80.06	9.64	61.50	9.53	63.54	7.72	5.03^{*}	First＞Fourth Third＞Fifth
Score-based Method	65.06	9.45	77.45	5.84	76.12	8.45	64.02	12.19	59.97	17.17	5.47^{**}	Second＞Fifth Third＞Fifth Second＞Fourth Third＞Fourth

Note. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

Table 2 shows that content-based filtering produced significant differences ($p < .05$), with the first ($M = 87.13$), second ($M = 83.71$), and third courses ($M = 78.72$) being higher than the fifth course ($M = 43.27$). Item-based collaborative filtering produced significant differences ($p < .05$), with the first ($M = 79.45$), second ($M = 71.09$), and third courses (M = 71.10) being higher than the fifth course (M = 56.74). User-based collaborative filtering produced significant differences ($p < .05$), with the first course (M = 69.51) being higher than the fourth course (M = 61.50) and the third course (M = 80.06) being higher than the fifth course (M = 63.54). The score-based method produced significant differences ($p < .05$), with the second (M = 77.45) and third courses (M = 76.12) being higher than the fifth course (M = 59.97) and the second (M = 77.45) and third courses (M = 76.12) being higher than the fourth course (M = 64.02). Overall, the students were more interested in the courses at the top of the recommendation lists.

5.3. The degree of interest in the recommended courses is affected by students' internal and external motivations for taking a course

The Mann-Whitney U nonparametric test is used in this section. The data follow a normal distribution (skewness between .29 and 1.57; kurtosis between -.09 and 2.39). Table 3 shows that the proportion of students with autonomous motivation (M = 44.49%~51.17%) was higher than that of students with extrinsic informational motivation (M = 18.89%~20.59%; $p < .05$) across the five recommendation methods. The results indicated that most students choose courses according to their plans, interests, or needs.

Table 3. A difference analysis of the students' motivation of course-taking in five recommendation methods

Recommendation methods	Autonomous motivation		Extrinsic informational motivation		p	Multiple comparison
	M	SD	M	SD		
Content-Based Filtering	49.21	16.90	19.15	15.25	$.00^{***}$	AM>EIM
Popularity-Based	46.29	16.35	19.96	16.35	$.00^{***}$	AM>EIM
Item-Based Collaborative Filtering	51.17	23.66	19.57	17.52	$.00^{***}$	AM>EIM
User-Based Collaborative Filtering	45.24	20.06	18.89	15.26	$.00^{***}$	AM>EIM
Score-Based	44.49	21.18	20.59	15.84	$.00^{***}$	AM>EIM

Note. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$. Autonomous Motivation = AM, Extrinsic Informational Motivation = EIM.

5.4. An analysis of the different degrees of interest in the courses recommended to a student using five recommendation methods and the degree of course suitability for the student

The Kruskal-Wallis nonparametric test is used in this section. The data followed a normal distribution (skewness between -1.01 and 2.09; kurtosis between -1.48 and 3.46). Table 4 shows that the students' interest matched between 60 and 70% of the course recommendation lists across the five recommendation methods, while there were no significant differences in the parameters according to the Kruskal-Wallis test ($p > .05$). Further analysis of the degree of alignment between student interest and the lists generated by the five recommendation methods showed that there were statistically significant differences in the parameters by the Kruskal-Wallis test ($p > .05$), and the results were the same for the degree of course list suitability for students. The post hoc comparisons showed that students thought that the results of the content-based filtering (M = 67.40) were more in line with their preferences and needs than other methods (Table 1).

Table 4. A difference analysis between the degrees of interest and suitability for student

Recommendation methods	Degree of interest (N = 46)				Degree of suitability (N = 46)			
	M	SD	p	Multiple comparison	M	SD	p	Multiple comparison
Content-based Filtering	70.14	21.58	.59	-	67.40	47.40	$.00^{***}$	CBF>PB CBF>IBCF CBF>UBCF CBF>SB
Popularity-based Method	66.03	25.55			32.60	47.40		
Item-based Collaborative Filtering	64.63	25.33			30.40	46.52		
User-based Collaborative Filtering	65.88	23.12			26.10	44.40		
Score-based Method	62.39	25.99			28.30	45.52		

Note. $^{***}p < .001$. Content-based Filtering = CBF, Popularity-based Method = PB, Item-based Collaborative Filtering = IBCF, User-based Collaborative Filtering = UBCF, Score-based Method = SB.

6. Discussion

In developing the PHCRS, we used ROC and NDCG to evaluate the system's accuracy. After the students used the PHCRS, the course fields that were only score-based showed obvious differences in the data analysis. This indicates that the students are less interested in the recommended courses when the list produced is not divided by field. In contrast, the students are more interested in their optional courses when the fields are divided in the recommendation list. The results partially support Hypothesis 1, which indicates that if the PHCRS considers the fields that the students are interested in, the recommendation accuracy increases. We also found that for all five recommendation methods, the students were more interested in the courses at the top of the list. This aligns with Babad (2001), who believes that students care most about informativeness, lecture quality and potential value for their future careers when selecting their first course. The courses selected toward the end of their education tend to be easier courses. While the results support Hypothesis 2, they also validate the appropriateness of the course order produced by the PHCRS.

When students referred to the course list provided by the five recommended methods, 44.49 to 51.17% of students chose courses based on autonomous motivation, which aligned well with the study list based on their interests and course content. Additionally, 18.89 to 20.59% of students chose courses based on extrinsic informational motivation, which caused them to consider how easy the course is to pass or earn a high grade in, the style of the lecturer or whether their colleagues are taking the same course. The results support Hypothesis 3 that students choose courses based on internal motivation and after considering their own interests (Barth, 2008; Wolbring & Treischl, 2016). Finally, approximately 60 to 70% of the students were interested in the course lists recommended to them by all five recommendation methods, and 67.40% of students said that content-based filtering produced the best results. Thus partially supporting Hypothesis 4. This indicates that most students choose courses on the basis of the course content, which is in line with previous relevant research that has concluded that content-based filtering is best suited to meet students' needs and are also most frequently used in course recommendation systems because it considers the characteristics of every course using syllabus details to provide course recommendations similar to students' previous courses (Apaza et al., 2014; Esteban et al., 2020).

7. Conclusion

This research applied five recommendation methods to build a PHCRS: content-based filtering, a popularity-based method, item-based collaborative filtering, user-based collaborative filtering, and a score-based method. Compared to recommendation systems built based on only one of these methods, our system is more suitable for fulfilling the diverse educational needs of students. We also built a labeling process that transfers text from course syllabi into a database, and a classification rule for information such as the field and objectives of the course and the knowledge, skills and perspectives students encounter or learn. Future studies can use the text database to enhance their course classification accuracy with text mining approaches. This database can also be a reference for other schools when developing a recommendation system for their electrical and computer engineering departments. After the system was built, to enhance the efficiency and make it more interesting for the students to use, the recommendation system website was coordinated with the school course selection website to assist students in selecting their courses before the start of every semester based on their personal needs. To help avoid students blindly selecting courses, we integrated the past course selection data of the student to calculate the ratio of autonomously selected courses to the courses selected using the recommendation list. When students log into the website, their course selection characteristics are automatically shown (Figure 2). Finally, we evaluated the performance of the PHCRS using recommendation metrics and questionnaires. The NDCG of the five recommendation methods is higher than .85, especially for user-based collaborative filtering (which had an NDCG of .96). The ROC of item-based collaborative filtering achieved .82. The results showed that the PHCRS can accurately predict students' needs and recommend suitable courses. In the questionnaires, we evaluated the effectiveness of the PHCRS on the basis of students' major, course selection order, and course selection motivations and the accordance between the recommended courses and the actual course selection. Overall, approximately 70% of the students were interested in the course list recommended by the PHCRS, which would shows that the system can guide students in choosing courses in their major and saves them time in choosing courses outside their major.

References

Aguilar, J., Valdiviezo-Díaz, P., & Riofrio, G. (2017). A General framework for intelligent recommender systems. *Applied Computing Informatics*, *13*, 147-160. https://doi.org/10.1016/j.aci.2016.08.002

Aher, S. B., & Lobo, L. M. R. J. (2013). Combination of machine learning algorithms for recommendation of courses in e-learning system based on historical data. *Knowledge-Based Systems*, *51*, 1-14. http://dx.doi.org/10.1016/j.knosys.2013.04.015

Apache Friends. (2021, June). *XAMPP Apache + MariaDB + PHP + Perl*. https://www.apachefriends.org/zh_tw/index.html

Apaza, R. G., Cervantes, E. V., Quispe, L. C., & Luna, J. O. (2014, September). *Online courses recommendation based on LDA* [Paper presentation]. Information Management and Big Data, Cusco, Peru. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.664.8706&rep=rep1&type=pdf

Archer, J. J., & Cooper, S. (1998). *Counselling and mental health services on campus: A Handbook of contemporary practices and challenges.* Jossey-Bass Inc.

Babad, E. (2001). Students' course selection: Differential considerations for first and last course. *Research in Higher Education*, *42*, 469-492. https://doi.org/10.1023/A:1011058926613

Babad, E., Avni-Babad, D., & Rosenthal, R. (2004). Prediction of students' evaluations from brief instances of professors' nonverbal behavior in defined instructional situations. *Social Psychology of Education*, *7*, 3-33. https://doi.org/10.1023/B:SPOE.0000010672.97522.c5

Babad, E., & Tayeb, A. (2003). Experimental analysis of students' course selection. *British Journal of Educational Psychology*, *73*, 373-393. https://doi.org/10.1348/000709903322275894

Barth, M. M. (2008). Deciphering student evaluations of teaching: A Factor analysis approach. *Journal of Education for Business*, *84*(1), 40-46. https://doi.org/10.3200/JOEB.84.1.40-46

Begs, J. M., Bentham, J. H., & Tyler, S. (2008). Distinguishing the factors influencing college students' choice of major. *College Student Journal, 4*(22), 381-394.

Bendakir, N., & Aïmeur, E. (2006). Using association rules for course recommendation. *American Association for Artificial Intelligence*, *3*, 1-10.

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, *12*(4), 331-370. https://doi.org/10.1023/A:1021240730564

Çano, E., & Morisio, M. (2017). Hybrid recommender systems: A Systematic literature review. *Intelligent Data Analysis, 21*(6), 1487-1524. https://doi.org/10.3233/IDA-163209

Chang, H. T., Wang, L. C., Lin, C. Y., Tseng, F. C., & Jheng, W. B. (2020, October). *LINE study secretary: AI-based personalized learning assistance system* [Paper presentation]. 2020 Conference on Innovation and Design (CID), Taipei, Taiwan. https://id100.chihlee.edu.tw/p/412-1025-3920.php

Cuseo, J. (2003). *Academic advisement and student retention: Empirical connections and systemic interventions*. National Academic Advising Association. Retrieved from http://www.uwc.edu/admin-istration/academic-affairs/esfy/cuseo/Academic%20Advisement%20and%20Student%20Retention.doc

Dwivedi, S., & Roshni V. S., D. K. (2017, August). *Recommender system for big data in education* [Paper presentation]. 5th National Conference on E-Learning & E-Learning Technologies, Hyderabad, India. https://doi.org/10.1109/ELELTECH.2017.8074993

Esteban, A., Zafra, A., & Romero, C. (2020). Helping university students to choose elective courses by using a hybrid multi-criteria recommendation system with genetic optimization. *Knowledge-Based Systems*, *194*(22), 1-14. https://doi.org/10.1016/j.knosys.2019.105385

Farzan, R., & Brusilovsky, P. (2006, June). *Social navigation support in a course recommendation system* [Paper presentation]. International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, Dublin, Ireland. https://link.springer.com/chapter/10.1007%2F1176801211

Gordon, V. N. (2007). *The Undecided college student: An Academic and career advising challenge*. Charles C Thomas Pub Ltd.

Han, J. W., Jo, J. C., Ji, H. S., & Lim, H. S. (2016). A Collaborative recommender system for learning courses considering the relevance of a learner's learning skills. *Cluster Computing*, *19*, 2273-2284. https://doi.org/10.1007/s10586-016-0670-x

Herlocker, J., Konstan, J., & Riedl, J. (2000). Explaining collaborative filtering recommendations. *2000 ACM conference on Computer supported cooperative work* (pp. 241–250). https://doi.org/10.1145/358916.358995

Hou, Y., Zhou, P., Xu, J., & Wu, D. O. (2018, April). *Course recommendation of MOOC with big data support: A Contextual online learning approach* [Paper presentation]. IEEE Conference on Computer Communications Workshops, Honolulu, HI. https://ieeexplore.ieee.org/document/8406936

Huang, L., Wang, C. D., Chao, H. Y., Lai, J. H., & Yu, P. S. (2019). A Score prediction approach for optional course recommendation via cross-user-domain collaborative filtering. *IEEE Access*, *7*, 19550-19563. https://doi.org/10.1109/ACCESS.2019.2897979

Iatrellis, O., Kameas, A., & Fitsilis, P. (2017). Academic advising systems: A Systematic literature review of empirical evidence. *Education Science, 7*(4), 1-17. https://doi.org/10.3390/educsci7040090

Kongsakun, K., Fung, C. C., & Chanakul, T. (2010). Developing an intelligent recommendation system for a private university in Thailand. *Issues in Information Systems*, *11*(1), 467-472.

Kurniadi, D., Abdurachman, E., Warnars, H. L. H. S., & Suparta, W. (2019). A Proposed framework in an intelligent recommender system for the college student. *Journal of Physics: Conference Series*, *1402*, 1-7. https://doi.org/10.1088/1742-6596/1402/6/066100

Lang, C. (2010). Happenstance and compromise: A Gendered analysis of students' computing degree course selection. *Computer Science Education*, *20*(4), 317-345. https://doi.org/10.1080/08993408.2010.527699

Lee, Y. M., & Sun, S. H. (2010). The Relationship between autonomous motivation of course-taking and learning engagement on college students. *Journal of Research in Education Sciences, 55*(1), 155-182. https://doi.org/10.3966/2073753X2010035501006

Mondal, B., Patra, O., Mishra, S., & Patra, P. (2020, March). *A Course recommendation system based on grades* [Paper presentation]. IEEE International Conference on Computer Science, Engineering and Applications (ICCSEA), Gunupur, India. https://ieeexplore.ieee.org/document/9132845

Obeid, C., Lahoud, I., El Khoury, H., & Champin, P. A. (2018). Ontology-based recommender system in higher education. *WWW'18: Companion Proceedings of the The Web Conference 2018* (pp. 1031–1034). https://doi.org/10.1145/3184558.3191533

Pardos, Z. A., & Jiang, W. (2020). Designing for serendipity in a university course recommendation system. *LAK'20: Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (pp. 350–359). https://doi.org/10.1145/3375462.3375524

Perera, A. T., & Pratheesh, P. (2018). Factors influence on student course selection: A Case of trincomalee campus. *International Journal of Innovative Science and Research Technology, 3*(5), 579-581.

Perugini, S., Goncalves, M. A., & Fox, E. A. (2004). Recommender systems research: A Connection-centric survey. *Journal of Intelligent Information Systems*, *23*(2), 107-143. https://doi.org/10.1023/B:JIIS.0000039532.05533.99

Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, *40*(3), 56-58. https://doi.org/10.1145/245108.245121

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *3*(1), 12-27. https://doi.org/10.1002/widm.1075

Salehudin, N. B., Kahtan, H., Adbulgabber, M. A., & Al-bashiri, H. (2019). A Proposed course recommender model based on collaborative filtering for course registration. *International Journal of Advanced Computer Science and Applications*, *10*(11), 162-168. https://doi.org/10.14569/IJACSA.2019.0101122

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. *WWW '01: Proceedings of the 10th international conference on World Wide Web* (pp. 285–295). https://doi.org/10.1145/371920.372071

Sawarkar, N., Raghuwanshi, M. M., & Singh, K. R. (2018). Intelligent recommendation system for higher education. *International Journal on Future Revolution in Computer Science & Communication Engineering*, *4*(4), 311-320.

Schafer, J. B., Konstan, J., & Riedl, J. (1999). *Recommender systems in e-commerce*. *EC '99: Proceedings of the 1st ACM conference on Electronic Commerce* (pp. 158–166). https://doi.org/10.1145/336992.337035

Tuckman, B. W. (1999). *Conducting educational research* (5th ed.). Sea Harbor Drive, Orlando: Harcourt Brace College Publishers.

Urata, U., & Takano, A. (2003). Construction of a student counseling system: Collaboration between psychology and college education. *Japanese Journal of Educational Psychology*, *51*, 205-217. https://doi.org/10.5926/jjep1953.51.2_205

Wang, L. C., Lin, J. Y., & Chang, H. T. (2020, June). *A Personalized course recommendation system for college students* [Paper presentation]. Association for Institutional Research Forum, New Orleans, LA. https://www.airweb.org/collaborate-learn/air-forum

Wolbring, T., & Treischl, E. (2016). Selection bias in students' evaluation of teaching. *Research in higher education*, *57*, 51-71. https://doi.org/10.1007/s11162-015-9378-7

WordPress.com. (2021). *WordPress*. https://wordpress.com/

Xiao, J., Wang, M., Jiang, B., & Li, J. (2018). A Personalized recommendation system with combinational algorithm for online learning. *Journal of Ambient Intelligence and Humanized Computing*, *9*, 667-677. https://doi.org/10.1007/s12652-017-0466-8.

Xu, J., Xing, T., & Schaar, M. van der. (2016). Personalized course sequence recommendations. *IEEE Transactions on Signal Processing, 64*(20), 5340-5352. https://doi.org/10.1109/TSP.2016.2595495

Yao, C. B. (2017). Constructing a user-friendly and smart ubiquitous personalized learning environment by using a context-aware mechanism. *IEEE Transactions on Learning Technologies, 10*(1), 104–114. https://doi.org/10.1109/TLT.2015.2487977

Zare-ee, A., & Shekarey, A. (2010). The Effects of social familial and personal factors on students' course selection in Iranian technical schools. *Procedia Social and Behavioral Sciences, 9*, 295-298. https://doi.org/10.1016/j.sbspro.2010.12.153

Zhang, H. R., Min, F., He, X., & Xu, Y. Y. (2015). A Hybrid recommender system based on user-recommender interaction. *Mathematical Problems in Engineering*, *2015*, 1-12. https://doi.org/10.1155/2015/145636

Zweig, M., Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: A Fundamental evaluation tool in clinical medicine. *Clinical Chemistry, 39*(8), 561-577. https://doi.org/10.1093/clinchem/39.4.561

Shen, J., Shvonski, A., Cui, T., & Wamba, S. F. (2022). Editorial Note: Understanding and Bridging Gap in Multi-mode Digital Learning during Post-Pandemic Recovery. *Educational Technology & Society*, *25* (1), 75-77.

Editorial Note: Understanding and Bridging Gap in Multi-mode Digital Learning during Post-Pandemic Recovery

Jun Shen[1*], Alex Shvonski[2], Tingru Cui[3] and Samuel Fosso Wamba[4]
[1]University of Wollongong, Australia // [2]Massachusetts Institute of Technology, USA // [3]University of Melbourne, Australia // [4]Toulouse Business School, France // jshen@uow.edu.au // shvonski@mit.edu // tingru.cui@unimelb.edu.au // s.fosso-wamba@tbs-education.fr
[*]Corresponding author

ABSTRACT: COVID-19 pandemic had changed the world-wide education landscape as the whole society is adapting to the "new normal." We orgainised a special issue collecting research papers that shed insights on how teaching and learning designs will be affected, and how novel educational technologies will help in a fast post-pandemic recovery. 26 papers were received but only 11 papers were finally selected to publish, after two rounds of rigorous reviews. This editorial note discusses the background, quality management and thematic topic groups of the papers.

Keywords: Multi-mode learning, Post-pandemic, Pedagogical, Learning behavior

1. Introduction

At the doorstep of the third decade in the 21st century, fast-growing computing technologies boost the adoption of diverse devices and applications in the educational area, which dazzle instructors and learners. The outbreak of COVID-19 since the beginning of 2020 made all learning activities online in many countries and territories, which adopted social distancing approaches to contain spread of virus. Unfortunately, many teachers and students have felt overwhelmed by such a drastic change in learning behaviour, though digital learning had been around for decades, including big efforts put in MOOC movements in different sectors. Quick adoption of digital means of learning and teaching might quickly fade out when the lifestyle get back to normal. It is thought-provoking to know if the unexpected pandemic brings the yet-to-come education evolution earlier.

Although being still demanding, the post-pandemic recovery is on the agenda. How the educational sector can stand with the contingency and bounce back stronger with insights gained during the pandemic pose interest. It pictures how the post-pandemic human development and learning looks like, allowing it to potentially shift from just content dissemination to augmenting relationships with teachers, personalization, and independence.

Evaluating the effectiveness and knowing in which environments the advanced technologies work better, and improving learning activities from both the students' and instructors' perspective are critical for the next generation delivery of the learning content. Given their comparative novelty, to what extent instructors and learners can accept and get accommodated to them sustain the ongoing update and development of new technologies. There are huge challenges ahead for understanding and bridging the gap in implementation of multi-mode digital learning over the coming decade.

2. Paper solicitation and review

From a timely standpoint, we had invited submissions reporting research studies on the development and application of advanced learning technologies on multi-mode learning, and understand their insights for education in post-pandemic recovery from the pedagogical and practical perspective. Accepted papers are anticipated to provide comprehensive results collected from empirical data and the corresponding analysis to consolidate research validity.

We fortunately received 26 submissions from many different countries and territories including Australia, China (incl. Hong Kong), India, Indonesia, Italy, Korea, Malaysia, Pakistan, Spain, Taiwan, Thailand, Turkey, United Kingdom and United State. We are pleased to noticed some highly cited researchers and top research universities in digital learning fields also contributed their latest work to be considered in this special issue.

After initial check, we desk rejected 2 papers due to scope discrepancies. Then we sent the remainder 24 papers for reviews. We are grateful to have nearly 60 internationally acclaimed academics from more than 20 countries helping us out in this process. Every paper was reviewed by 3-4 reviewers. Most paper were rigorously checked by reviewers in two rounds of reviews to ensure the best papers were accepted and the published articles in this special issue to represent the highest quality suitable for this prestigious journal. Therein, we only selected 11 papers for final production, which we would briefly introduce next.

3. Accepted articles and topics

In this special issue, the articles cover different groups of topics.

3.1. Multi-mode course design during COVID-19 pandemic

Firstly, Pérez-Marín et al. (2022) are interested in applying VARK model in multi-mode of teaching HCI. Then, McLaughlan's (2022) article focuses on teacher training programs which are evaluated through contribution analysis. Huang et al. (26) has implemented a video-facilitated transdisciplinary STEM curriculum during the pandemic. With a slightly different angle from other papers in this collection, Rof et al. (2022) reflects deeply on how the learning value proposition of higher education institutions has been affected by the COVID-19 accelerated digital transformation.

3.2. Observational studies on teaching and learning behaviour

The paper co-authored by cross institutional team Hong et al. (2022) investigates the ineffectiveness of online learning related to cognitive and affective factors by paying attention to students' mind states during COVID-19 lockdown. Lin et al. (2022b) reports their comparative study on students' performance and attention to Khan-style video lecture (VL) and online practice (OP) group. Next, Sun et al. (2022) presents their empirical observation how primary school students respond to robots used in life sciences teaching. Guo et al. (2022) studies the effectiveness of 3D design in developing students' spatial ability.

3.3. Novel technology mediated teaching and learning for faster pandemic recovery

Yong et al. (2022) proposes to apply AI technologies in improving art courses' teaching, by highlighting the challenges with regard to online sharing of learning resources. On the other hand, Lin et al. (2022a) further applies natural language learning technologies in recommending micro learning materials. Zhai et al. (2022) observes effects of multi-mode stimuli on students' metacognition through novel eye tracking techniques.

4. Summary

Same as all sectors in the society during the pandemic, the whole process in organising this special issue took longer time and more coordinated efforts from the guest editorial team in order to collect high quality reviews. We are indebted to all reviewers' unselfishness and all authors' patience. Everybody has dedicated a lot of time and efforts to putting together this special issue.

Finally, we hope this collection of articles will contribute to the literature for the related research communities.

Reference

Guo, S., Wang, X., Deng, W., Hong, J., Wang, J., & Wu, Y. (2022). Whose spatial ability benefits from learning with 3D design? From the perspective of learning analysis. *Educational Technology & Society, 25*(1).

Hong, J.-C., Liu, X., Cao, W., Tai, K.-H., & Zhao, L. (2022). Effects of self-efficacy and online learning mind state on learning ineffectiveness during the COVID-19 lockdown. *Educational Technology & Society, 25*(1).

Huang, B., Jong, M. S.-Y., & Chai, C. S. (2022). Design and implementation of a video-facilitated transdisciplinary STEM curriculum in the context of Covid-19 pandemic. *Educational Technology & Society, 25*(1).

Lin, J., Sun, G., Beydoun, G., & Li, L. (2022a). Applying machine translation and language modelling strategies for the recommendation task of micro learning service. *Educational Technology & Society, 25*(1).

Lin, C.-H., Wu, W.-H., & Lee, T.-N. (2022b). Using an online learning platform to show students' achievements and attention in the video lecture and online practice learning environments. *Educational Technology & Society, 25*(1).

McLaughlan, T. (2021). Chai Chats: An Online teacher-training program of observation and social connectedness evaluated via contribution analysis. *Educational Technology & Society, 25*(1).

Pérez-Marín, D., Paredes-Velasco, M., & Pizarro, C. (2022). Multi-mode digital teaching and learning of human-computer interaction using the VARK model during COVID-19. *Educational Technology & Society, 25*(1).

Rof, A., Bikfalvi, A., & Marques, P. (2022). Pandemic-accelerated digital transformation of a born digital higher education institution: Towards a customized multimode learning strategy. *Educational Technology & Society, 25*(1).

Sun, Y., Chang, C.-H., & Chiang, F.-K. (2022). When life science meets educational robotics: A Study of students' problem solving process in a primary school. *Educational Technology & Society, 25*(1).

Yong, B., Jiang, X., Lin, J., Sun, G., & Zhou, Q. (2022). Online practical deep learning education: Using collective intelligence from a resource sharing perspective. *Educational Technology & Society, 25*(1).

Zhai, X., Chu, X., Meng, N., Wang, M., Spector, M., Tsai, C.-C., & Liu, H. (2022). The Effect of multi-mode stimuli of feedforward and eye tracking on metacognition— An Exploratory study using digital dictionaries. *Educational Technology & Society, 25*(1).

Multi-mode Digital Teaching and Learning of Human-Computer Interaction (HCI) using the VARK Model during COVID-19

Diana Pérez-Marín[1*], Maximiliano Paredes-Velasco[1] and Celeste Pizarro[2]
[1]Escuela Técnica Superior de Ingeniería Informática, Universidad Rey Juan Carlos, Spain // [2]Escuela Superior de Ciencias Experimentales y Tecnología, Universidad Rey Juan Carlos, Spain // diana.perez@urjc.es // maximiliano.paredes@urjc.es // celeste.pizarro@urjc.es
[*]Corresponding author

ABSTRACT: In this paper, a multi-mode digital teaching approach is proposed based on the use of the VARK (Visual, Aural, Read/Write, Kinaesthetic) model where students have different styles (one or more) that improve their learning (face-to-face and online). Our research question is on the effectiveness of this approach in terms of learning efficacy and students' satisfaction. An experiment with 41 students has been carried out for five months to answer the research question and to provide a first validation of using VARK for multi-mode digital HCI teaching. During the experiment, the theoretical sessions were given through videoconference using Microsoft Teams and with the support of Moodle. In the practical sessions, students had to create a software prototype following a User-Centred Design with a real client. For this, they used Discord to collaborate in their groups, Teams to ask questions to teachers and PowerPoint and Genially to present their work online to the class through a Teams videoconference. A regression model has been provided to predict the VARK indicated by the questionnaire to each student with a prediction success of nearly 77%. Using the VARK multi-mode digital teaching approach has proved valid, and effective and beneficial in the teaching of HCI with a significant improvement in the learning scores and satisfaction levels of the students even with respect to pre-COVID-19 where the teaching was face-to-face.

Keywords: Multi-mode digital teaching, VARK, Human-Computer Interaction, COVID-19

1. Introduction

Human-Computer Interaction is the discipline related to the design, evaluation, and implementation of interactive systems for human use and the study of the related phenomena (Card et al., 1983). The goal is to remove barriers in the dialogue between users and systems. The idea is that interfaces should be designed to facilitate users in accomplishing their goals, regardless of whether the goals are to complete a report or to play a videogame. HCI has undergone significant research in recent decades since the use of technological devices is no longer restricted to people with technical knowledge but to everybody (Stephanidis et al., 2019).

In our university, HCI has traditionally been taught face-to-face (F2F) in all Computer Engineering degrees. Students in groups of 40-100 were sat in a classroom following the traditional lecture combined with exercises in class and practical activities in the computer lab. A review of how HCI was taught at other universities also revealed that this is the common approach in F2F universities, at least in Spain (Pérez-Marín, 2018).

However, due to COVID, teaching of the subject had to move online in March 2020, even in F2F universities. Several approaches have been tried to hold online seminars (e.g., Seminar Series in HCI at Carnegie Mellon University in 2020), proposing recommended readings in HCI (e.g., Human-Computer Interaction course at the University of Cambridge), videoconferences using Microsoft Teams or Blackboard Collaborate, and/or uploading multimedia and teaching notes at virtual campus using Moodle or other e-learning platforms. A combination of these approaches following a multi-mode teaching and learning approach considering the students' preferences may be key to significant learning and satisfaction among students even during COVID-19. Our proposal is to use the VARK model (Fleming & Mills, 1992) to digitally teach HCI in a multi-mode format to university students. The research question is how effective a multi-mode digital approach using the VARK model can be in terms of learning efficacy and student satisfaction. An experiment with 41 HCI university students has been carried out to answer the research question and test the hypothesis.

2. Related work

This section provides the context of this research with a review of the theoretical background provided in Section 2.1. and more technical background in Section 2.2.

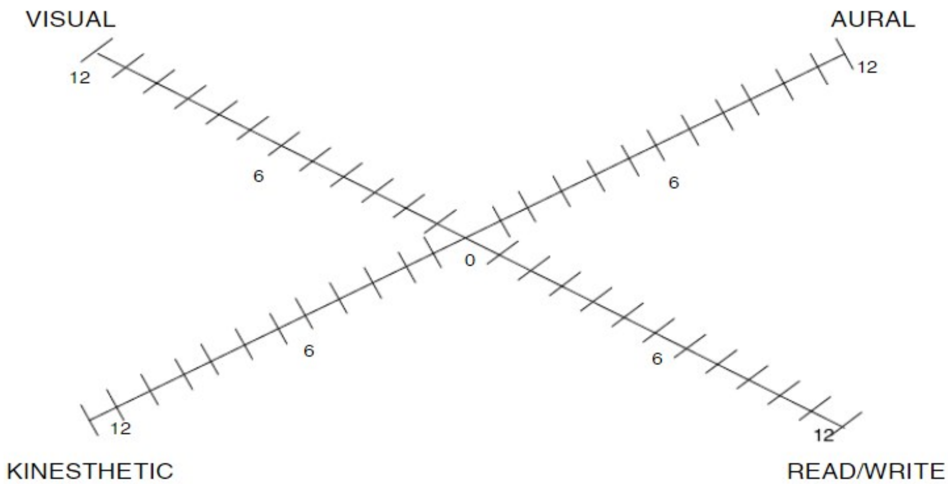
2.1. Theoretical background

Multi-mode teaching can be defined as the combination of multiple modes of knowledge representation such as oral and written language, visual, gestural, tactile, and spatial representations (Jewitt, 2008; Cope & Kalantzis, 2009). Much research has been focused on exploring how to design effective multi-mode digital teaching experiences (Bell et al., 2010; Cope & Kalantzis, 2009). A decade later, multi-mode digital teaching seems to have become key in overcoming the difficulties of teaching and learning during COVID-19 (Oyedotun, 2020; Teele et al., 2020). Multi-mode teaching facilitates imparting information, enacting collaborative learning, and preparing students for exploring concepts (Papageorgiou & Lameras, 2017). Moreover, combined digital learning technologies can help students develop technical and non-technical skills such as creativity, capacity of innovation and problem solving (Philippe et al., 2020).

According to Bakar (2007), there are at least five variables that should also be considered when creating an effective instructional model: active student involvement, attracting interest and attention, raising their motivation, individual principle, and displays used in lessons. When using online courses and students' learning styles are reflected in their design, their learning efficacy is higher (Lee & Choi, 2011; El-Bishouty et al., 2019). However, when searching IEEE Xplore, Elsevier, Web of Science and SCOPUS for papers on multimodal learning platforms and experiences using learning styles, much has been written focusing on styles such as those proposed by Felder and Silverman (1988), i.e., active/reflective, sensing/intuitive, visual/verbal, sequential/global, but there is a gap on systems and platforms using other models that could also benefit students in online courses such as the VARK model (Fleming & Mills, 1992).

The VARK (Visual, Aural, Read/Write, Kinaesthetic) model is based on the idea of empowering students by finding out their sensory preferences and adjusting their study methods accordingly. The core idea, in the words of Fleming (1995), is "in observing the best of teachers apparently there is no single best way to teach but teachers who cater for the different needs of students by using a variety of teaching approaches are rewarded with improved learning (p. 1)." There are, however, no hard boundaries between the styles and students can have one or more styles combined. To find out students' sensory preferences, a questionnaire with multiple-choice questions was created in English (VARK, 2021), and later translated into 14 languages (Fleming & Baume, 2006). It currently has 16 questions with four possible answers per question. Students are told to choose the option that best matches their perception, but if they do not feel that one single answer is the perfect match, they can choose more than one option. Students can leave a question unanswered if they feel the question does not apply to them. The minimum value for each preference is 0, and the maximum is 12 (Hawk & Shah, 2007) as shown in Figure 1.

Figure 1. VARK learning model (Source: Hawk & Shah, 2007)



A strong preference in a style can be identified by a score four or five points higher than any other. A difference below two points between preferences is not enough. A void—or a score of one—on a mode would suggest that it is a weak preference for that student. Fleming (2001) reported that about 41% of students taking the questionnaire online had single style preferences, 27% two preferences, 9% three preferences and 21% four preferences. Table 1 gathers activities that accommodate VARK preferences. The fourth preferences according to Fleming (1995) relate to different types of content and activity: visual students (V) prefer graphs, charts, and flow diagrams; aural students (A) prefer sounds and audio; read and write students (R and W) prefer documents and notes; kinaesthetic students (K) prefer experiences and samples; and multi-mode students (MM) prefer several possibilities.

Table 1. Suggested activities for VARK preferences (Source: Fleming, 2001)

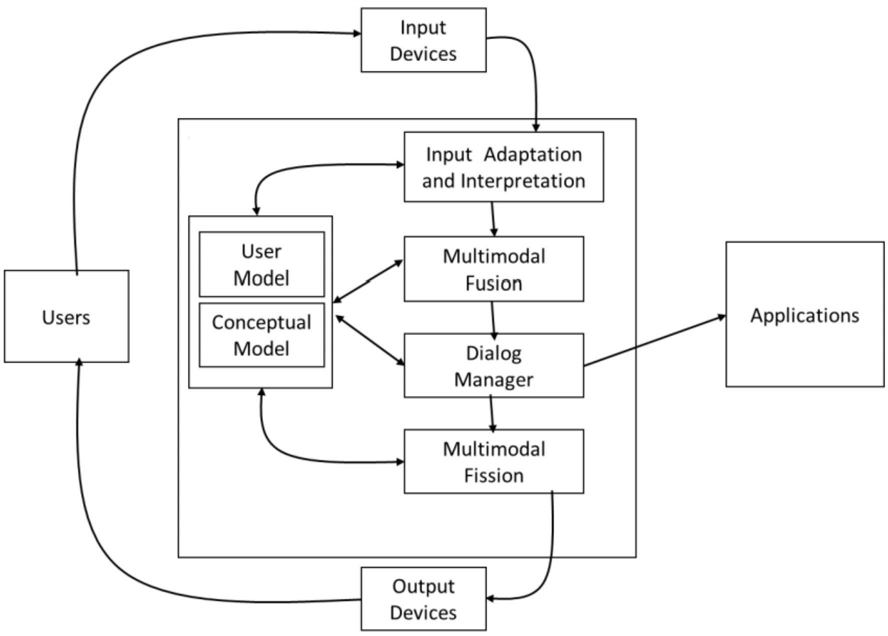
Visual	Aural	Read/Write	Kinaesthetic
Diagrams	Debates, arguments	Books, texts	Real-life examples
Graphs	Discussions	Handouts	Examples
Colours	Conversations	Reading	Guest lecturers
Charts	Audio tapes	Written feedback	Demonstrations
Written texts	Video+audio	Note taking	Physical activity
Different fonts	Seminars	Essays	Constructing
Spatial arrangement	Music	Multiple choice	Role-play
Designs	Drama	Bibliographies	Working models

2.2. Practical experiments

Fleming (2001) presented results revealing higher student performance when studying according to the VARK preferences indicated by the questionnaire. Fleming & Baume (2006) reported over 180,000 people having used VARK online between mid-March and mid-September 2006, and an attempt of validation although reliability values are not provided. Instead, a warning is given that the questionnaire is not to be used as a diagnosis tool, and explained by the creator as a stimulus to reflect upon. Experiments have been carried out to find out whether students found that the VARK preferences as indicated by the model are what they expected (Espinoza-Poves et al., 2019) and the pedagogical implications of the VARK model and how it can generally be used for online teaching (González, 2012; Hussain, 2017; Prithishkumar, 2014).

After performing an experiment with 92 Nurse Education students with a single group pre-post study, Alkhasawneh et al. (2008) found a significant increase in their grades after their VARK preferences has been taken into account. Their underlying teaching methodology was Problem-Based Learning. Moazeni and Pourmohammadi (2013) provided an automatic student modelling approach for distance education to optimise the teaching strategy to align instruction with students' learning styles using the VARK model. However, they did not provide any platform to implement and validate their approach. Similarly, Stojanova et al. (2017) highlighted the benefits of using the VARK model to teach a Data Structure and Algorithms course, which is the closest work to this paper together with Díaz et al. (2018) as it is also in the Computer Science domain (although not in HCI where literature is scarce). They used Moodle, presentations and animations using Java Applets and/or Flash and videos from YouTube. Neither a teaching methodology nor a framework to implement their approach is given.

Figure 2. Overview of the multimodal framework proposed by Vidakis (2017)



Lee and Kim (2016) proposed the Multimodal Teaching Learning Model (MTLM) based on providing the teacher with feedback and scaffolding to increase the interest of students. The use of technology using synchronous environments in small groups is considered beneficial in MTLM, although the use of learning preferences was not investigated. They carried out an experiment to test MTLM in the language-learning domain. A significant improvement in students' knowledge was found. Moreover, 54.9% of the students reported feeling a stronger bond with their teacher, and 62.9% of the students reported a stronger bond with their classmates, even in distance education.

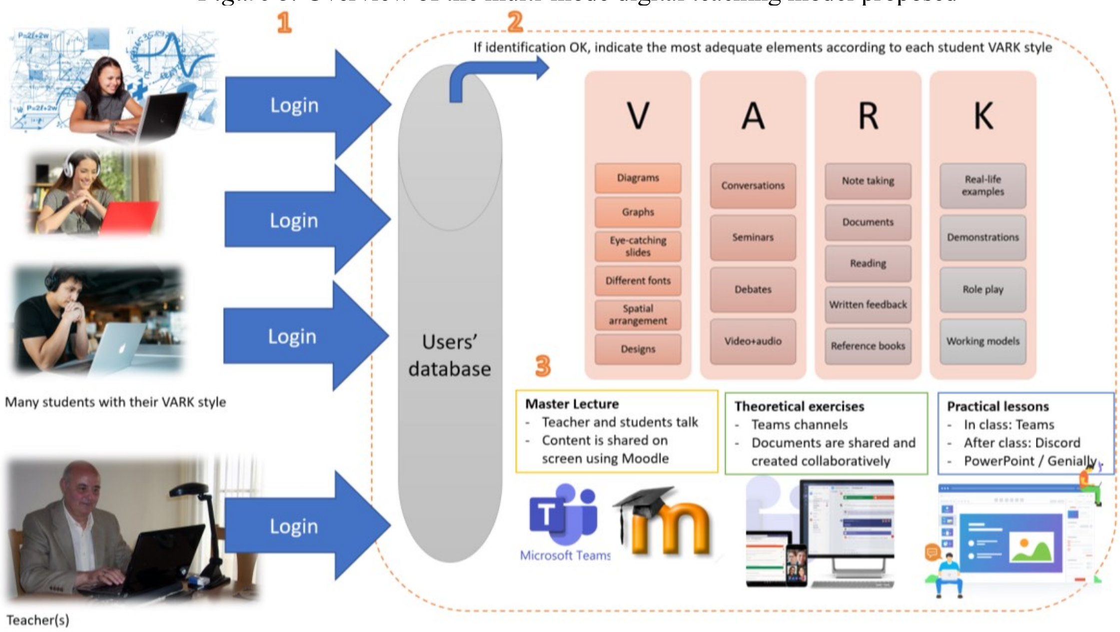
Vidakis (2017) described a multimodal framework to enable the deployment of more interaction modalities between students and online educational systems than just speech and touch. However, the paper does not mention the benefits of also considering the modality in the course contents. Figure 2 shows an overview of the multimodal framework proposed. No experiment is provided in which the platform has been implemented and used by students.

Finally, Díaz et al. (2018) highlighted how modified VARK styles and adaptive learning materials benefit both students and teachers. They proposed a platform to support the learning of object orientation with VARK learning styles. However, their focus was more on the creation of adaptive material as they called it, than in the computer system itself.

3. Proposal: Multi-mode digital teaching and learning of HCI using the VARK model

A shift from the traditional large-group teacher-centric teaching to a student-centric multi-mode digital approach is necessary (Prithishkumar, 2014). Models such as VARK are useful in highlighting the diversity of preferences among students and that there is no one teaching solution that matches all their preferences and domains. In this section, a proposal for multi-mode teaching and learning of theoretical and practical online lessons for HCI is described together with the architecture of a platform to support it given the lack of such platforms in the literature (see Section 2).

Figure 3. Overview of the multi-mode digital teaching model proposed



In F2F teaching, traditional theoretical lessons are associated with Master lectures. The reason why it continues to be used could be that it is cheap—one teacher with a relatively high number of students. However, with a Master lecture, it is not possible to manage the diversity of preferences among students; the lesson is just the same for all students with the possible consequence of bored students, unable to fully understand the lessons (Spanish Education Ministry, 2006). They should therefore be combined with exercises and practical lessons. Moreover, as already indicated by Lee and Kim (2016), the use of synchronous communications systems such as WhatsApp or Teams creating small groups is beneficial. Collaborative learning is the chosen underlying pedagogic model to mitigate the isolation feeling that students may have in distance education. Figure 3 overviews the proposal for multi-mode teaching and learning of HCI covering VARK styles.

As can be seen in Figure 3, initially (step 1) all students and teachers must login into the platform to identify themselves. This is a key step as the style of each student can be saved in the database. Thus, the system can indicate (step 2) which content elements from the VARK styles are the most appropriate for the teacher according to students' styles. For the Master lecture (step 3), teachers can use videoconference software such as Teams. As they talk, aural (A) students would benefit from listening to the lesson and visual (V) students would appreciate the use of eye-catching slides so that the knowledge is not only spoken but also written with a good design (e.g., using a template with different fonts, diagrams, graphs, and some spatial arrangement that leaves spaces between paragraphs). Read-Write (R) students will be taking notes during the lesson, and they will appreciate an activity such as writing a document to sum up the ideas of the lesson and to get some written feedback. A list of reference books is also suitable for R students to improve their learning. Kinaesthetic (K) students will appreciate the use of real-life examples, demonstrations, and working models as well as role-play exercises to understand the contents of the lesson better.

Creating channels in Teams is recommended to group students so that they can discuss the lesson and complete the exercises with another 4-5 students (Walter, 1983; Lee & Kim, 2016). Activities solely with the whole group are not advisable, as some students would not participate. For instance, as a possible exercise for the theoretical lesson for all VARK students would be that when teaching User-Centred Design (UCD, Lorés, 2002; Abras et al., 2004), after looking at the slides (V), listening to the teacher (A), note taking (R) and providing some detailed examples (K), students in their channels could think of one interactive system (e.g., a new videogame) to design following a UCD. V students would start creating the interface's diagram, A students would talk about the design and would understand the UCD better by listening to their classmates (not only the teacher) creating debate about what to do, R students would create a document to upload to the channel so that the teacher can check that their debate and thinking is correct (otherwise some spoken and written feedback should be provided to address the specific mistakes detected) and K students would have an example to understand the general UCD process.

For practical activities in the computer lab, the use of videoconference software such as Teams is also recommended for teachers to explain the activities to complete. Practical activities are a necessary complement to theoretical lessons (De Miguel, 2006) and the content should be at the same pace. Practical activities should follow the same recommendations previously provided for the Master lecture to address all different VARK preferences and channels should also be created for students to complete the practical activities for the reasons explained above. For instance, a possible practical project for all VARK students would be that when completing their practical work, they had images, diagrams, graphs, charts of a UCD (V), they could talk in their groups about the phases of a UCD and discuss them (A), they could read books describing the phases of a UCD (R) and write a report about that, and have a real example of creating a prototype following a UCD with a real client (K). Students should also present their practical work to their classmates (the whole class, not just their group). The goal would be for V students in other channels to see designs other than their own, A students would listen to different conversations about the topic, R students would have more documents to read and write about, and K students would have more examples to improve all their understanding. Students should be given the freedom to choose the presentation software to use. They may prefer traditional PowerPoint, just sharing their desktop, and talking about their activities from the documents created (no need to create new ones), or newer possibilities such as Genially. Students should have these new possibilities such as multi-layer content and animated templates explained to them to enhance interactivity and integrate knowledge.

V students will appreciate the aesthetic design of the templates in any presentation software. They will use them to make them easier to write their content. Moreover, they will not need to prepare different slides, because it is possible to create pop-ups that are shown as needed by clicking on them. A students will also appreciate the possibility of adding music and audio recordings to the presentation. R students will be reading the contents of the lessons and reference books to write the content of the presentation. Finally, K students will populate the contents of the presentation with many examples and specific cases to illustrate their points.

4. Research method

The research method used is a mixed quantitative and qualitative experimental research study with control and test groups. The theoretical justification for this method is the need to test a hypothesis and answer the research questions associated with a practical experiment in the field of software systems to gather data to perform both a descriptive and inferential statistical analysis, and to complement this quantitative data with the qualitative data provided by the users of the systems from answers to questionnaires (Goundar, 2013; Lorés, 2002).

Following an adaptation of the guidelines to report experiments in Engineering domains written by Jedlitschka and Pfahl (2005), and Wohlin et al. (2012) this section is structured as follows: 4.1. Goal, 4.2. Participants and Context, 4.3. Experimental Materials, 4.4 Procedure, and 4.5. Variables.

4.1. Goal

The goal of the experiment is to validate the proposed digital multi-mode teaching model using the VARK model as described in Section 3 in terms of learning efficacy, students' satisfaction levels and reliability of the VARK preferences. Although VARK is not a diagnosis tool, as explained in Section 2, we believe that it should be confirmed whether the VARK preference provided to students in the questionnaire reflects their personal preferences to really help them act accordingly.

The research questions together with the hypotheses are:

RQ1. Are the learning scores of students following the digital multi-mode teaching and learning model using VARK for HCI higher than the learning scores of students following traditional F2F lessons?

H1. Students following the digital multi-mode teaching and learning model using VARK will achieve higher scores than students following traditional F2F lessons.

RQ2. How satisfied are teachers and students following the digital multi-mode teaching and learning model using VARK?

H2. Teachers and students following the digital multi-mode teaching and learning model using VARK will be satisfied, and they will prefer it to F2F lessons.

RQ3. Are preferences provided by the VARK questionnaire valid?

H3. The preferences provided by VARK questionnaire are valid, as they will be supported by the answers from students to a preferences questionnaire.

4.2. Participants and context

The experiment was conducted in the first semester of the 2020/2021 academic year from September 2020 to January 2021 with 41 students enrolled in the third year of the Videogame design degree at the university. Of the students, 72% are between 20-22 years old, 20% are between 24-26 years old, and 8% are older than 26 years old. The split by sex is 80% men and 20% women. They have a high level of digital competence, enjoy using technology, and have a positive attitude towards its general use.

There were 44 students in the 2019-2020 academic year (control group) with a similar distribution of age and sex, digital competence level, and positive attitude towards technology. The main difference is that the Videogame design degree is a F2F degree at our university, thus the Master lectures and practical activities were F2F during that first semester from September 2019 to January 2020. For those students, therefore, the teaching and learning was F2F. However, due to COVID-19, it was agreed that Master lectures and practical activities should move online from March 2020. Therefore, students in the 2020-2021 academic year (test group) followed the online digital multi-mode teaching and learning model using VARK.

Both courses were taught by the same two teachers, one man and one woman who are both experts in HCI. All students were voluntarily asked to participate in the experiment. No increase of score or reward was given to any student. The motivation provided was focused on the goal of the experiment being to improve the teaching of HCI to more students, and that they learn about different modes of learning, and eventually get recommendations about their study preferences to improve their learning.

4.3. Experimental materials

All educational materials were created by the subject's teachers. The content of the subject in both the 2019/2020 and 2020/2021 academic years was the same, with the difference that the presentations were given in class in 2019/2020, and shared through the Teams videoconference system and uploaded to the digital campus hosted in

Moodle in 2020/2021. To cover all VARK preferences, the content was a set of slides with an eye-catching design, diagrams, written text, references, documents, and videos, as well as documents, videos, and external links to references in books and on websites.

4.4. Procedure

Students in the 2019/2020 academic year attended F2F theoretical lessons with the same content as in the 2020/2021 academic year. The exercises and practical activities were the same. The only difference was that the VARK multi-mode digital approach was followed in 2020/2021, and they were F2F in 2019/2020 without considering VARK preferences. The step-by-step procedure for the control group was:

(1) Lessons started in September 2019.
(2) For each week, students attended F2F classes on:
    2.1 Tuesdays (2 hours): Master lectures with theoretical exercises
    2.2 Fridays (2 hours): practical lessons
(3) Lessons finished in January 2020, and students took their final exam.

The step-by-step procedure for test group (using the multimodal methodology with VARK) was:

(1) Lessons started in September 2020.
(2) For each week, students attended classes online on:
    2.1 Tuesdays (2 hours): Master lectures with theoretical exercises
    2.2 Fridays (2 hours): practical lessons
(3) Lessons finished in January 2021, and:
    3.1 Students took their final exam.
    3.2 Students were asked to complete the VARK questionnaire.
    3.3 Students were asked to complete an online questionnaire about their experience.

Table 2. Final questionnaire for the students

Question	Possible answers	Measure
To see a Teams videoconference	1 (minimum) – 5 (maximum)	Preferences
To see a Blackboard videoconference	1 (minimum) – 5 (maximum)	Preferences
To talk about the exercises in groups	1 (minimum) – 5 (maximum)	Preferences
To write an individual report	1 (minimum) – 5 (maximum)	Preferences
To write a report in groups	1 (minimum) – 5 (maximum)	Preferences
To individually present the content	1 (minimum) – 5 (maximum)	Preferences
To present the content in groups	1 (minimum) – 5 (maximum)	Preferences
To create a video of the content on my own	1 (minimum) – 5 (maximum)	Preferences
To create a video of the content with my group	1 (minimum) – 5 (maximum)	Preferences
To read the slides	1 (minimum) – 5 (maximum)	Preferences
To take notes on paper	1 (minimum) – 5 (maximum)	Preferences
To take notes digitally	1 (minimum) – 5 (maximum)	Preferences
Do you think that presenting your practical work to the other students helped your learning of the subject?	Yes/No	Opinion
If face-to-face lessons were possible, would you prefer to attend face-to-face lessons instead of online lessons?	Yes/No	Satisfaction
Do you think that you would have learnt more following a face-to-face teaching approach?	Yes/No	Opinion
Do you think that you would have been happier if the teaching were face-to-face?	Yes/No	Satisfaction
Do you think that creating a prototype with a real client has helped your learning of the subject?	Yes/No	Opinion
Any other comment?	Free text	Opinion & Satisfaction

The Master lectures, theoretical exercises and practical work had the same structure, similar difficulty, content, weight in the final score (60%), and were evaluated on the same scale from 0 (minimum) to 10 (maximum). Samples of the practical activities have been published by Pérez-Marín (2018). The final exam was F2F in both courses with the same structure: three theoretical questions about the same concepts and with the same difficulty in both academic years with a maximum score of three points; one question to draw a prototype with a maximum score of four points; and one final question to write a report on assessing the usability of a videogame that they could freely choose with a maximum score of three points. The exam was completed individually without any help from the internet or reference books. It accounted for 40% of the final grade in both academic years, and the scale of the exam was also the same from 0 (minimum) to 10 (maximum). A sample exam can be found in Pérez-Marín (2018).

Only students in the test group in 2020/2021, when applying the VARK model as described in Section 3, were also asked to complete two additional questionnaires individually and online at the end of the course: (1) the Spanish translation of the VARK questionnaire (Sámano-Galindo & Preciado-Delgado, 2007); (2) a questionnaire to mark their preferences, and their satisfaction regarding the multi-mode digital teaching on a Likert scale (from 1-minimum to 5-maximum); to answer Yes/No to some opinion and satisfaction questions, and any other comment they may have, to give them the opportunity of freely expressing themselves. The questionnaire was not anonymous because the intention was to relate the values gathered with the results of the VARK questionnaire. In any event, no name or personal information was asked for because it would be contrary to Spanish law. A code was therefore created from their practical group number and their position on the list of group members. Without the list of groups, therefore, it was impossible to identify the students. Table 2 shows the questions with their possible answers and what they measure.

4.5. Variables

The dependent variables of the study were related, firstly, to learning efficacy, measured by scores obtained for the students at the end of the experiment, named *Score*. These ratings are divided into two groups, those who had F2F teaching and those who had digital multi-mode teaching; the factor variable named *Group* shows these differences. Secondly, a categorical dependent variable named *VARKL* collects the results of the VARK questionnaire into the four preferences described in Section 3: A, R, K and MM. Additionally, a group of 12 ordinal variables scaled from 1 to 5 collect the results gathered in the preferences questions described in Section 4.4. They will be called x_i in relation to the i-th question. Table 3 summarises the variables used in the experiment.

Table 3. Summary of variables

Aspect	Type	Variable	Name
Learning HCI	DV	Scores	Score
	IV	Use of the digital VARK multi-model	Group
Preferences	DV	VARK questionnaire	VARKL
	IV	i-th question in questionnaire for the students' preferences (i = 1, ... , 12)	x_i

Note. DV: Dependent variable, IV: Independent variable, name, and description.

5. Results

5.1. Learning efficacy

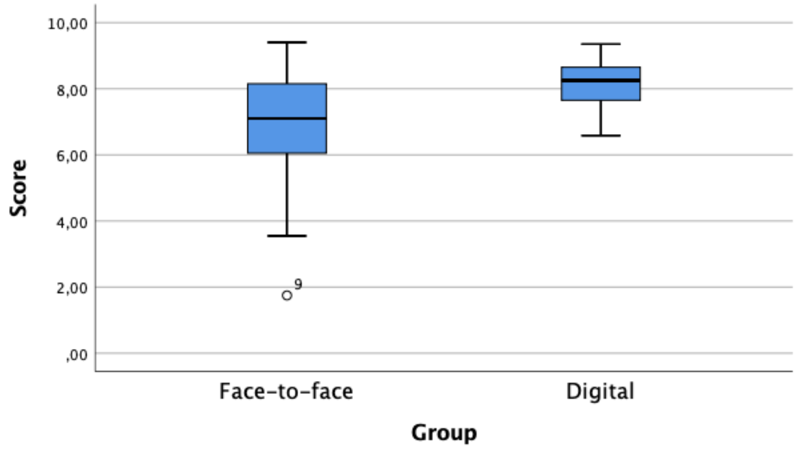
Table 4 shows the mean and median values (more representative than mean in asymmetric distributions) and standard deviations for the scores in the final exam of the control and test groups. In the F2F control group, the mean of the marks is 6.90, a value somewhat lower than the median, with a standard deviation of 1.71. In the test multi-mode digital group, the mean is more than one point higher at 8.10, again somewhat lower than its median of 8.25. For this group, in addition, the dispersion of the scores is much smaller, the difference with the previous case being more than one point: 0.69. Figure 4 shows a graphical summary of these data as boxplots.

After checking the normality of the data in both groups (Shapiro-Wilk test with $p > .05$), the t-test for independent groups was chosen. Table 4 shows the significant increase of the scores in the digital multi-mode group, with $p < .001$.

Table 4. Descriptive analysis of the scores in the final exam for both groups

Group	*N*	Mean	Median	*SD*	*t*-test	*df*	*p*-value
Control (F2F)	44	6.90	7.10	1.71	4.272	57.78	< .001
Test (Multimodal)	41	8.10	8.25	0.69			

Figure 4. Boxplots for the scores in the final exam of the F2F control and digital test groups



5.2. Students' satisfaction

Regarding RQ2, students seem satisfied with the multi-mode digital teaching with answers such as "Following this approach I think I learn a lot, so I like that", "I like that all theoretical lessons are practiced later with a real client" or "I like these lessons very much! Thanks!" to the final open question. In total, 84% of students considered that explaining the practical work to other classmates helped them as shown in Figure 5 and in general, they do not think that they would have learnt more in a F2F lesson (only 35% students answered that they would learn more in a F2F lesson as shown in Figure 6). In total, 56% of students thought that they would have enjoyed the F2F lessons more (see Figure 7).

When the HCI teachers were asked about their satisfaction with using the VARK multi-mode digital approach compared to classical teaching, they both agreed that they were more satisfied because they could combine multiple resources and saw that their students understood the lessons faster and better.

Figure 5. Satisfaction with the practical work (84% yes, 16% no)

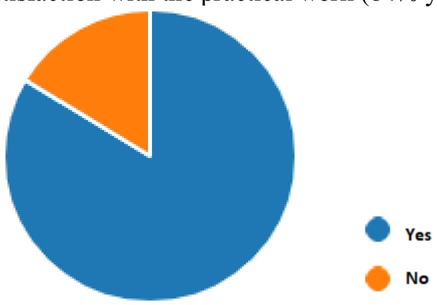


Figure 6. Opinion about whether they thought they would learn more with F2F teaching (35% yes, 65% no)

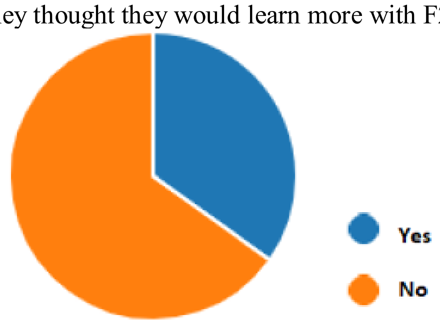


Figure 7. Answer to the question: do you think that you would be happier if the teaching were F2F? (47% yes, 53% no)

5.3. VARK validation

To see the possible relationship between the dependent variable VARKL and the 12 independent variables of the preferences questions, a multinomial logistic regression model was chosen, since the output variable is a categorical variable, and the predictor variables are ordinal variables. The absence of multicollinearity between the variables is first verified, with the FIV value being under 10 in all independent variables, and with a tolerance greater than 0.1 in all cases.

The model fitting information provided a $\chi^2(48) = 85.59$ ($p = .001$), i.e., the full model predicts the dependent variable better than the intercept-only model alone. Moreover, Pearson $\chi^2(108) = 43.643$ and Deviance $\chi^2(108) = 39.937$, both with $p > .005$. Pseudo Nagelkerke $R^2 = 0.626$, a medium value that measures the degree of improvement in the fit of the log-likelihood model with respect to the model without independent variables. The model gives an overall correct classification percentage of 76.9%.

6. Discussion

6.1. Theoretical contributions

The results of this study revealed that students following the VARK multi-mode online approach significantly improved learning outcomes with respect to traditional F2F students (scores 8.10 vs 6.90 respectively). Therefore, RQ1 is positive and H1 is accepted. As reviewed in Section 2, there are many studies regarding multi-mode teaching effectiveness. Yeh (2018) investigated students' perceptions with respect to their knowledge level in English as a foreign language, where students had to produce a digital video, reflective essays and PowerPoint slides employing multiple modalities and formats. Yeh discovered that students perceived that the process based on creating compositions, oral presentations and video editing improved their multiliteracies to learn the target language. Other works have not focused on the student's perception, but on self-assessment processes to determine multimodal approach effectiveness in an educative context, finding that a multi-mode blended learning model produces significant improvements in several language-learning skills (Chen, 2018; Lee & Kim, 2016). Santana-Mancilla et al., (2019) found that the use of teaching methods based on games in HCI education, which has multi-mode interaction, provides students with important skills in this area, such as involving users, task-centred system design, models of human behaviour, creativity and metaphors, and graphical screen design.

However, the previous multimodal models do not always consider students' preferences with an underlying pedagogic model. The multimodal teaching learning model proposed in the article is based on VARK, which groups several learning preferences and a collaborative learning approach together. Therefore, the paper contributes with a model that guides learning activities and instructional development considering different learning preferences and combines pedagogical methods with digital tools, the most appropriate way for HCI education.

The paper also has two more contributions regarding innovations in learning effectiveness. Firstly, most experience in previous research regarding learning effectiveness was mainly developed in a literacy education context. In this paper, the learning experience was developed in computer science learning, specifically Human-Computer Interaction, which has been poorly researched regarding multi-mode digital teaching approaches. Secondly, most research in the literature applies subjective assessment based on students' perception to measure learning effectiveness, whereas this study develops objective assessment using exam scores.

6.2. Theoretical implications

Regarding to the use of VARK in HCI education context, the results show that hypothesis H3 must be accepted. RQ3 is positive, as it has been possible to predict the preferences provided by the VARK questionnaire as the output value of a regression model with 77% success.

To the best knowledge of the authors, the use of VARK in HCI learning is rare. However, some work in learning programming, which is an educative context close to HCI can be found such as Stojanova et al. (2017), who applied VARK to learn data structures and algorithms with discussion tools using Moodle LMS. Additionally, they integrated visualisations of algorithms with VisuAlgo (see http://visualgo.net/) and discovered that the use of VARK improved interest among students and kept their attention in class. Díaz et al. (2018) carried out an experience with industrial engineering students of object-oriented programming courses. Students had access to an adaptive e-learning platform that proposed different learning contents and tasks according to VARK style. Díaz et al. (2018) discovered that the predominant VARK style preferences of engineering students were kinaesthetic and aural and that there were no visual style students. This finding is aligned with the results found in this paper, where the predominant preferences of the HCI students enrolled in the Videogame design degree were kinaesthetic, and only one student was visual.

6.3. Practical implications

Learning outcomes do not only depend on a pedagogy approach. There are emotional and affective factors that may increase learning efficacy (Lin et al., 2016; Urquiza-Fuentes & Paredes-Velasco, 2017). Students' satisfaction experienced with multi-mode digital learning has been analysed in this study too with the answer to RQ2 being that both teachers and students were very satisfied. In total, 84% of students were satisfied with the practical class model used and only 35% of students considered F2F to be more efficient. Thus, H2 is accepted. These results are aligned with other studies on HCI multimodal education. For instance, high satisfaction experienced by participants in the experience reported is similar to positive emotions experienced by other students that worked with multimodal interaction approach, "Regarding if they enjoyed learning using computer games, 100% of the students enjoyed the course" and "100% said that the knowledge acquired would have been lower [if the teaching process were not related to video game design]" (Santana-Mancilla et al., 2019, p. 9).

The authors consider that this phenomenon is explained by multi-mode environments influencing students' beliefs and perception about their skills and knowledge during the learning. Banzato and Coin (2019) carried out an experience where students had to develop multimodal narrative learning activities through gestural/mime languages, drawings, oral presentations and compositions and they found that a multimodal approach influenced students' self-efficacy beliefs about their narrative skills. Santana-Mancilla et al. (2019) stated that multimodal interaction in HCI education promotes students having a positive perception on the efficacy of the use design for practical works. In addition, the use of multimodal information spaces, rich in digital and physical resources (Facebook discussion, downwards projection, tablets, etc.) raise students' satisfaction and motivation, and contribute to their engagement and collaboration in HCI learning (Vasiliou et al., 2013). These studies are aligned with the results of the experience reported in this paper, where 53% of students perceived the use of practical work as the best way to understand theoretical contents and Teams and Blackboard video-collaborative platforms as the best tools to learn content digitally. The practical contribution of this paper is the digital ecosystem defined by the multimodal teaching learning model, which facilities applying multimodality in a practical way in HCI education and improving students' engagement and satisfaction.

6.4. Limitations

Regarding possible threats to this study, the following issues have been considered: all questionnaires were applied at the end of the experience to avoid influencing the results; repeat students did not participate in the experience so the pre-pandemic (traditional F2F) and post-pandemic (multi-mode online learning) groups were different; and, although the control group was from last year and test group from this year, learning contents, tasks and teachers were the same for both groups, pre and post-pandemic, with only the teaching and learning methodology changing. However, some threats to validity are recognised in the experiment presented (Campbell & Stanley, 1963; Cook & Campbell, 1979; Shadish et al., 2002):

- **Internal validity:** the scenario in which the test group was must be considered; in most cases, the student's home. This can be an advantage, as many of them operate in an environment that they consider safer than university. Others, on the other hand, may have worse digital media available, etc.

- **External validity**: since the groups are not created randomly—it depends on the students enrolled in each academic year—there is no certainty that the sample is representative of the general population.
- **Construct validity**: an important part of applying the VARK method lies in the use of new methodologies. The use of audiovisual methods during the development of the subject is a novelty in some activities, such as debates, seminars, etc. According to Bracht and Glass (1968) there is a certain enthusiasm when there is innovation, and this can contribute to success.
- **Conclusion validity:** This is concerned with sources of random error and with the appropriate use of statistics and statistical tests (Cook & Campbell, 1979). They are also called SCV. The work presents a broad statistical study, which involves, firstly, a descriptive analysis of the data, combining descriptive and graphic techniques. Subsequently, an inferential study is carried out, in which the necessary conditions have been previously verified. Still, type 1 errors (incorrect rejection of the null hypothesis) and type 2 errors (not rejecting a false null hypothesis), although minimised, might be present.

7. Conclusions

The most relevant contribution to HCI education is a detailed and validated digital multi-mode teaching and learning approach using the VARK model. As indicated by Ioannou et al. (2015), this research contribution is particularly beneficial for HCI courses given that digital teaching is highly accepted by HCI students, provides satisfaction, and raises the acquisition level of HCI knowledge. The paper also contributes to multi-mode digital teaching and learning using the VARK model with a new validated approach supported by a framework of how to implement it that could guide other researchers and teachers to put it into practice in their lessons. It is particularly relevant as no similar framework has been found in the literature. Future research will focus on keeping investigating factors influencing the multimodal learning environment.

References

Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-centered design. In W. Bainbridge (Ed.), *Encyclopedia of Human-Computer Interaction* (pp. 445-456). Thousand Oaks, CA: Sage Publications

Alkhasawneh, I. M., Mrayyan, M. T., Docherty, C., Alashram, S., & Yousef, H. Y. (2008). Problem-based learning (PBL): assessing students' learning preferences using VARK. *Nurse education today*, 28(5), 572-579.

Bakar, A. (2007). Pengaruh kemahiran teknikal pensyarah politeknik terhadap proses pengajaran dan pembelajaran mengikut persepsi pelajar [The influence of technical skills of polytechnic lecturers on the teaching and learning process according to students' perceptions] (Unpublished master dissertation). Universiti Tun Hussein Onn Malaysia, Batu Pahat, Malaysia.

Banzato, M., & Coin, F. (2019). Self-Efficacy in multimodal narrative educational activities: Explorative study in a multicultural and multilingual italian primary school. *Media and Communication*, *7*, 148. https://doi.org/10.17645/mac.v7i2.1922

Bell, T., Urhahne, D., Schanze, S., & Ploetzner, R. (2010). Collaborative inquiry learning: Models, tools, and challenges. *International Journal of Science Education*, *32*(3), 349-377

Bracht, G. H., & Glass, G. V. (1968). The External validity of experiments, *American Education Research Journal*, *5*(4), 437- 474.

Campbell, D., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand-McNally.

Card, S. K., Newell, A., & Moran, T. P. (1983). *The Psychology of Human-Computer Interaction*. L. Erlbaum Associates

Carnegie Mellon University. (2020). *Seminar series in human-computer interaction*. https://www.hcii.cmu.edu/seminar-series

Chen, Z. (2018). Teaching effect of the multi-mode blended learning model from students' perceptions. In S. Cheung, J. Lam, K. Li, O. Au, W. Ma, & W. Ho (Eds.), *International Conference on Technology in Education* (pp. 188–199). Springer. https://doi.org/10.1007/978-981-13-0008-0_18

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin.

Cope, B., & Kalantzis, M. (2009). "Multiliteracies": New literacies, new learning. *Pedagogies: An International Journal*, *4*(3), 164-195.

De Miguel, M. (2006). Metodologías de Enseñanza Centrados en el Desarrollo de Competencias [Teaching Methodologies focused on the Development of Competencies]. Ediciones Universidad de Oviedo.

Díaz, F., Rubilar, T., Cubillos, C., & Silva, R. (2018). An Adaptive e-learning platform with VARK learning styles to support the learning of object orientation. *Proceedings of the 2018 IEEE World Engineering Education Conference (EDUNINE)* (pp. 6). https://doi.org/10.1109/EDUNINE.2018.8450990

El-Bishouty, M. M., Aldraiweesh, A., Alturki, U., Tortorella, R., Yang, J., Chang, T. W., & Graf, S. (2019). Use of Felder and Silverman learning style model for online course design. *Educational Technology Research and Development*, *67*(1), 161-177.

Espinoza-Poves, J. L., Miranda-Vilchez, W. A., & Chafloque-Céspedes, R. (2019). Los estilos de aprendizaje Vark en estudiantes universitarios de las escuelas de negocios [VARK learning styles in college students from business schools]. *Propósitos y Representaciones*, *7*(2), 384-414. http://dx.doi.org/10.20511/pyr2019.v7n2.254

Felder, R. M., & Silverman, L. K. (1988). Learning styles and teaching styles in engineering education. *Engineering Education*, *78*(7), 674–681.

Fleming, N., & Baume, D. (2006) Learning Styles Again: VARKing up the right tree! *Educational Developments*, *7*(4), 4-7.

Fleming, N. D., & Mills, C. (1992). Not Another Inventory, Rather a Catalyst for Reflection. *To Improve the Academy*, 11, 137-155.

Fleming, N. D. (2001). *Teaching and learning styles: VARK strategies*. New Zealand: Christchurch.

Fleming, N. D. (1995). I'm different; not dumb. Modes of presentation (VARK) in the tertiary classroom. In A. Zelmer (Ed.), Research and Development in Higher Education, *Proceedings of the 1995 Annual Conference of the Higher Education and Research Development Society of Australasia (HERDSA)* (Vol. 18, pp. 308- 313). http://www.vark-learn.com/wp-content/uploads/2014/08/different_not_dumb.pdf

González, B. (2012). El modelo VARK y el diseño de cursos en línea [The VARK model and online course design]. *Revista mexicana de bachillerato a distancia*, *4*(8). http://dx.doi.org/10.22201/cuaed.20074751e.2012.8.44282

Goundar, S. (2013). Research Methodology and Research Method. In S. Goundar (Ed.), *Cloud Computing* (pp. 2-43). Victoria University of Wellington.

Hawk, F., & Shah, A. (2007). Using learning style instruments to enhance student learning. *Decision Sciences Journal of Innovative Education*, *5*(1), 1-19. https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1540-4609.2007.00125.x

Hussain, I. (2017). Pedagogical implications of VARK model of learning. *Journal of Literature, Languages and Linguistics*, *38*, 33-37.

Ioannou, A., Vasiliou, C., Zaphiris, P., Arh, T., Klobučar, T., & Pipan, M. (2015). Creative multimodal learning environments and blended interaction for problem-based activity in HCI education. *TechTrends*, *59*, 47–56. https://doi.org/10.1007/s11528-015-0839-9

Jedlitschka, A., & Pfahl, D. (2005, November). *Reporting guidelines for controlled experiments in software engineering* [Paper presentation]. IEEE International Symposium on Empirical Software Engineering, Queensland, Australia.

Jewitt, C. (2008). Multimodality and literacy in school classrooms. *Review of Research in Education*, *32*(1), 241-267.

Lee, Y., & Choi, J. (2011). A Review of online course dropout research: Implications for practice and future research. *Educational Technology Research and Development*, *59*, 593–618. https://doi.org/10.1007/s1142 3-010-9177-y

Lee, K., & Kim B. (2016). Cross space: The Exploration of SNS-based writing activities in a multimodal learning environment. *Educational Technology & Society*, *19*(2), 57-76.

Lin, H., Su, S., Chao, C.-J., Hsieh, C.-Y., & Tsai, S.-C. (2016). Construction of multi-mode affective learning system: taking affective design as an example. *Educational Technology & Society*, *19*(2), 132-147.

Lorés, J. (2002). *La Interacción Persona-Ordenador* [Human-Computer Interaction]. Asociación Interacción Persona Ordenador.

Moazeni, S., & Pourmohammadi, H. (2013). Smart teaching quantitative topics through the VARK learning styles model. In 2013 *IEEE Integrated STEM Education Conference* (ISEC) (pp. 1-7). IEEE.

Oyedotun, T. D. (2020). Sudden change of pedagogy in education driven by COVID-19: Perspectives and evaluation from a developing country. *Research in Globalization*, 2. https://doi.org/10.1016/j.resglo.2020.100029

Papageorgiou, V., & Lameras, P. (2017). Multimodal Teaching and Learning with the Use of Technology: Meanings, Practices and Discourses. *14th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA)*, 133-140.

Pérez-Marín, D. (2018). *Proyecto Docente* [Teaching Project]. Publicaciones de la Universidad Rey Juan Carlos. http://hdl.handle.net/10115/17069

Philippe, S., Souchet, A.D., Lameras, P., Petridis, P., Caporal, J., Coldeboeuf, G., & Duzan, H. (2020). Multimodal teaching, learning and training in Virtual Reality: A Review and case study. *Virtual Reality & Intelligent Hardware*, *2*(5), 421-442. https://doi.org/10.1016/J.Vrih.2020.07.008

Prithishkumar, I. J. (2014). Understanding your student: Using the VARK model. *Journal of Postgraduate Medicine, 60*(2), 183-186. https://doi.org/10.4103/0022-3859.132337

Sámano-Galindo, R., & Preciado-Delgado, E. (2007). *Traducción del cuestionario VARK al español* [Translation of VARK questionnaire to Spanish]. https://vark-learn.com/el-cuestionario-vark/

Santana-Mancilla, P. C., Rodríguez-Ortiz, M. A., García-Ruiz, M. A., Gaytan-Lugo, L. S., Fajardo-Flores, S. B., & Contreras-Castillo, J. (2019). Teaching HCI skills in higher education through game design: A Study of students' perceptions. *Informatics*, *6*(2), 22. https://doi.org/10.3390/informatics6020022

Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002). *Experimental and quasi- experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Spanish Education Ministry. (2006). *Propuesta para la renovación de las metodologías educativas en la Universidad. Consejo de Coordinación Universitaria* [Proposal for the renewal of educational methodologies at the University. University Coordination Council]. Secretaría General Técnica. Subdirección General de Información y Publicaciones. http://publicaciones.administracion.es

Stephanidis, C., Salvendy, G., Antona, M., Chen, J., Dong, J., Duffy, V., Fang, X., Fidopiastis, C., Fragomeni, G., Fu, L.P., Guo, Y., Harris, D., Ioannou, A., Jeong, K., Konomi, S., Krömker, H., Kurosu, M., Lewis, J.R., Marcus, A., Meiselwitz, G., Moallem, A., Mori, H., Nah, F., Ntoa, S., Rau, P., Schmorrow, D., Siau, K., Streitz, N., Wang, W., Yamamoto, S., Zaphiris, P., & Zhou, J. (2019). Seven HCI grand challenges. *International Journal of Human–Computer Interaction*, *35*(14), 1229-1269. https://doi.org/10.1080/10447318.2019.1619259

Stojanova, A., Stojkovikj, N., Kocaleva, M., Zlatanovska, B., & Martinovska-Bande, C. (2017) Application of VARK learning model on "Data structures and algorithms" course. In *IEEE Global Engineering Education Conference (EDUCON)* (pp. 613-620). https://doi.org/10.1109/EDUCON.2017.7942909

Teele, A. A., Sindelar, A., Brown, D., Kane, D. A., Thatte, N., Williams, R. J., Gueverra, J., & Wolbrink, T. (2020). Online education in a hurry: Delivering pediatric graduate medical education during COVID-19. *Progress in Pediatric Cardiology*. https://doi.org/10.1016/j.ppedcard.2020.101320

Urquiza-Fuentes, J., & Paredes-Velasco, M. (2017). Investigating the effect of realistic projects on students' motivation, the case of Human-Computer interaction course. *Computers in Human Behavior*, *72*, 692–700. https://doi.org/10.1016/j.chb.2016.07.020

VARK. (2021). *On-line VARK questionnaire in English.* https://vark-learn.com/the-vark-questionnaire/ with the description of the modes at https://vark-learn.com/introduction-to-vark/the-vark-modalities/

Vasiliou, C., Ioannou, A., & Zaphiris, P. (2013). Technology enhanced PBL in HCI education: A Case study. In P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, & M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2013* (pp. 643–650). Springer. https://doi.org/10.1007/978-3-642-40498-6_56

Vidakis, N. (2017). A Multimodal interaction framework for blended learning. In *EAI Endorsed Transactions on Creative Technologies, 4*(10). https://doi.org/10.4108/eai.4-9-2017.153057

Walter, R. (1983). La realización de estudios de casos en educación. *Ética, teoría y procedimientos* [Conducting case studies in Education. Ethics, theory, and procedures]. In W. Dockrell & D. Hamilton (Eds.), *Nuevas reflexiones sobre la investigación educativa*. Narcea.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). *Experimentation in software engineering*. Springer Science & Business Media.

Yeh, H.-C. (2018). Exploring the perceived benefits of the process of multimodal video making in developing multiliteracies. *Language Learning and Technology*, *22*, 28–37.

Chai Chats: An Online Teacher-Training Program of Observation and Social Connectedness Evaluated via Contribution Analysis

Toni McLaughlan
Lancaster University, United Kingdom // t.mclaughlan@lancaster.ac.uk

ABSTRACT: This paper evaluates an international, online, content skills-based teacher education program sponsored by the U.S. State Department. The evaluation was designed using a RUFDATA framework (Saunders, 2000) to facilitate a complete, reflective assessment of the target program. Establishing causes-and-effects of the program's performance and data analysis involved adoption of a Contribution Analysis (Mayne, 2001; Mayne, 2008). Utilizing the six steps detailed in Mayne (2012), a credible contribution story emerged, highlighting strengths and weaknesses of transitioning teacher-training programs to virtual platforms. This evaluation has implications for teachers, teacher trainers, professionals planning similar programs particularly in developing regions, and individuals interested in how theory can be applied practically to impact continued teacher education processes. This paper contributes to knowledge as there are few formal evaluations of online international teacher education programs that facilitate observation of all aspects of a virtual course over an extended period of time and provide small-group engagement with course creators, especially with populations straddling the digital divide. It is also the first to conduct a theory-based evaluation of a U.S. English Language Specialist project despite the program's 1991 inception and current running rate of 150-200 projects annually worldwide (U.S. Department of State, 2021).

Keywords: Teacher training, Educational technology, Online, Developing regions, Contribution Analysis

1. Introduction

The integration of educational technology, especially resulting from Covid-19, has been growing for years. It is essential that educators are trained in effective teaching strategies involving technology and delivery of online instruction. One may question, then, whether using observation and practical applications of such tools may be an effective approach to teacher-training in these areas. This paper aims to evaluate an international, online, content skills-based teacher training program sponsored by India's Regional English Language Offices (RELO), part of the United States Bureau of Educational and Cultural Affairs (ECA), a division of the United States' Department of State.

This evaluation was first designed using the RUFDATA framework (Saunders, 2000) in order to facilitate the development of a complete and reflective assessment of the target program. To address remaining questions in analyzing the evidence gathered and establishing causes-and-effects of the program's performance, the evaluation adheres to a Contribution Analysis (Mayne, 2001; Mayne, 2008) approach and its six steps as detailed in Mayne (2012) in an effort to establish a strong, credible contribution story.

The paper begins with a situational analysis of the object of the evaluation, followed by a brief background of both RUFDATA and Contribution Analysis (CA). Subsequently, the methodology behind data collection in supporting the CA is outlined. The data analysis section then details each of CA's six steps, culminating in a revised contribution story. The paper concludes with discussion of the evaluation process, its limitations, and implications for future research, educators, and program developers.

1.1. Situational analysis of the object of the evaluation

1.1.1. Object of evaluation

The object of this evaluation is a professional development (PD) program for higher education (HE) instructors geared toward improving capabilities in the online teaching of academic English skills using educational technology throughout India. Though the program as a whole has multiple components, only the "Chai Chat" piece will be evaluated here. Participants observed a full 10-week virtual course that demonstrated the instruction of academic English reading or writing skills; each skill was taught for one hour weekly. This observation was done asynchronously; each lesson was taught and recorded through Zoom and shared with participants, who, while observing, completed weekly observation task forms (Wajnryb, 1992) that guided participants to focus on several highlighted aspects of a given lesson. These task forms were then submitted and reviewed by the RELO-selected English Language Specialist for each Chai Chat group which consisted of roughly 15-30 participants. Chai Chat groups met synchronously on Zoom with their Specialist for an hour weekly to discuss content, educational technology, and overall teaching methodology of the week's online lesson. During these meetings, instructors were encouraged to share ideas on how to best adapt lessons to their own contexts, such as larger/smaller groups, disadvantaged populations, academically at-risk students, asynchronous online teaching, or in-person instruction. They were encouraged to share instances of their own adoption of lessons' approaches, ask questions regarding incorporation of new technologies, share resources, and suggest preferred platforms for doing so. By the end of the program, instructors engaged in observing all elements of a full course: lessons, homework, rubrics, grading, online gradebooks, LMSs, and feedback to students.

1.1.2. Situational analysis

Institutions of higher education (HEIs) throughout India were suddenly faced with challenges of transitioning to distance learning with the emergence of Covid-19 (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2020). Closures of student and faculty residences often accompanied closures of campuses (Careers360, 2020), posing challenges of accessibility to internet connectivity and technological devices for both students and faculty. India's Ministry of Rural Development (2018) reports that only 47% of Indian households can reliably access electricity for 12+ hours a day. Furthermore, a mere 23.8% of Indian homes have internet access; 10.7% have a computer at home (Ministry of Statistics & Programme Implementation, 2018). Therefore, thriving in distance education is unlikely for the majority of tertiary students and faculty, and a challenge even for those having *some* accessibility capabilities. Additional studies of the region highlight a need for instructor training in online educational tools and strategies in order for web-based instruction to be effectual (Lakshmi & Agarwal, 2017); similarly, students themselves may lack digital literacy skills for educational purposes (Naresh et al., 2016).

2. Informing frameworks of the evaluation

2.1. RUFDATA

In designing a solid foundation for program evaluation, the meta-evaluative RUFDATA tool (Saunders, 2000) was adopted in directing the various reflexive questioning processes. The framework requires its users in initial planning stages to define key concepts comprising the RUFDATA acronym as they relate to the evaluation. The RUFDATA framework (Saunders, 2000) was applied as follows:

- *Reasons and purposes:* To provide value on the effectiveness of an intervening program developed to fulfill an unexpected yet substantial need in online instructional strategies in the Indian HE system
- *Uses:* By U.S.- and internationally-based employees of the ECA section of the U.S Public Affairs and Public Diplomacy division
- *Foci:* The "Chai Chai" component of a larger distance learning initiative
- *Data and evidence:* From individuals working with key stakeholders, program participants, their students, and program students in the form of surveys, program evaluations, and interviews
- *Audience:* A wide range of professionals in education, curriculum design, and training; the project's stakeholders
- *Timing:* Several weeks after program completion
- *Agency***:** RELO and its personnel under guidance of the evaluator, who also serves as the program's primary instructor and one of the five Specialists

2.2. Contribution analysis

Contribution Analysis (Mayne, 2001) is an evaluative approach that assists evaluators in addressing issues of cause-and-effect in assessing an intervening program's impacts. It is theory-based in that it involves the development of a *theory of change* (Weiss, 1995; Weiss, 1997) intended to model the expected changes of the program under evaluation, identifying various causal mechanisms that influence results of the program throughout (Patton, 2008). When an experimental evaluative design for inferring causality is not plausible, as is the case here, CA can be utilized in measuring a program's outcomes and impacts at various stages while providing explanations for why results do or do not occur.

3. Materials and methods

This program makes use of a mixed-methods design. Quantitative calculations present only descriptive data; no experimental element could be practically employed, and causation is instead addressed qualitatively through CA.

Throughout the program, participants were asked to complete observation task forms, adapted from Wajnryb (1992), who states that observation serves as a teacher training tool as educators gain skills of analyzing and interpreting while viewing others' techniques; this is then used for self-reflection. Observers completed the forms weekly, requiring them to record information from the lesson with a dual focus on a pedagogical element (i.e., classroom management, instructional language) and the content of the lesson's objective. This data informed Specialists on how to modify their Chai Chat meetings to be optimally relevant and guided the academic skills instructor in adapting the program's remaining lessons. For evaluation purposes, task form submissions were used primarily in reviewing attrition rates.

Three weeks following the program's end and into a new academic year, program evaluations were sent to participants via email. The program evaluations were developed and collected through Google Forms and included questions adapted from the Student Evaluation Quality Questionnaire (Marsh, 1982). The SEEQ is among the most utilized measures of course evaluation worldwide and is recognized for its consistent international reliability and validity, both attested over time (Coffey & Gibbs, 2001). Two weeks following review of program evaluation data, students of the academic skills courses were sent post-surveys inquiring about changes to their instructors' teaching habits using SurveyMonkey. This was done to corroborate participants' self-assessments on program evaluations. Similarly, Specialists received questionnaires via Google Forms regarding their participants' integration of program materials. Assumptions were further supported with data from student course evaluations. These evaluations were not initially intended for Chai Chat evaluation and were fully conducted under the direction of RELO, though review of its content reflects similar adaptation of the SEEQ (Marsh, 1982).

In addressing RELO stakeholders' program expectations, relevant data from Specialist questionnaires were reviewed, and semi-structured interviews were conducted with on-site RELO staff in India. Though representatives of English Language Programs with the U.S. Department of State are aware of and support the development of this evaluation and its possible publication, U.S.-based Embassy employees declined to provide information regarding in-house evaluation processes of individual Specialist projects. RELO stakeholders' goals mirror those listed on the ECA's website (Bureau of Educational and Cultural Affairs, n.d.).

4. Results

Using CA outlined in Mayne (2012), all six steps of the iterative analytical approach are detailed below. This includes a comprehensive Theory of Change (ToC), adapted from Mayne (2015).

4.1. Step 1: Set out the cause-effect issue to be addressed

In order to identify cause-effect issues related to the intervention, it is necessary to first clearly state the intended effects of the program. The primary goals of the intervention being evaluated are to:

- enhance instructors' effectiveness in teaching online and use of educational technologies,
- improve instructors' teaching efficacy by modeling a robust set of educational strategies in academic reading and writing in English,
- build public diplomacy through the promotion of English language programming, and
- foster mutual understanding between the U.S. and other nations through cultural exchange.

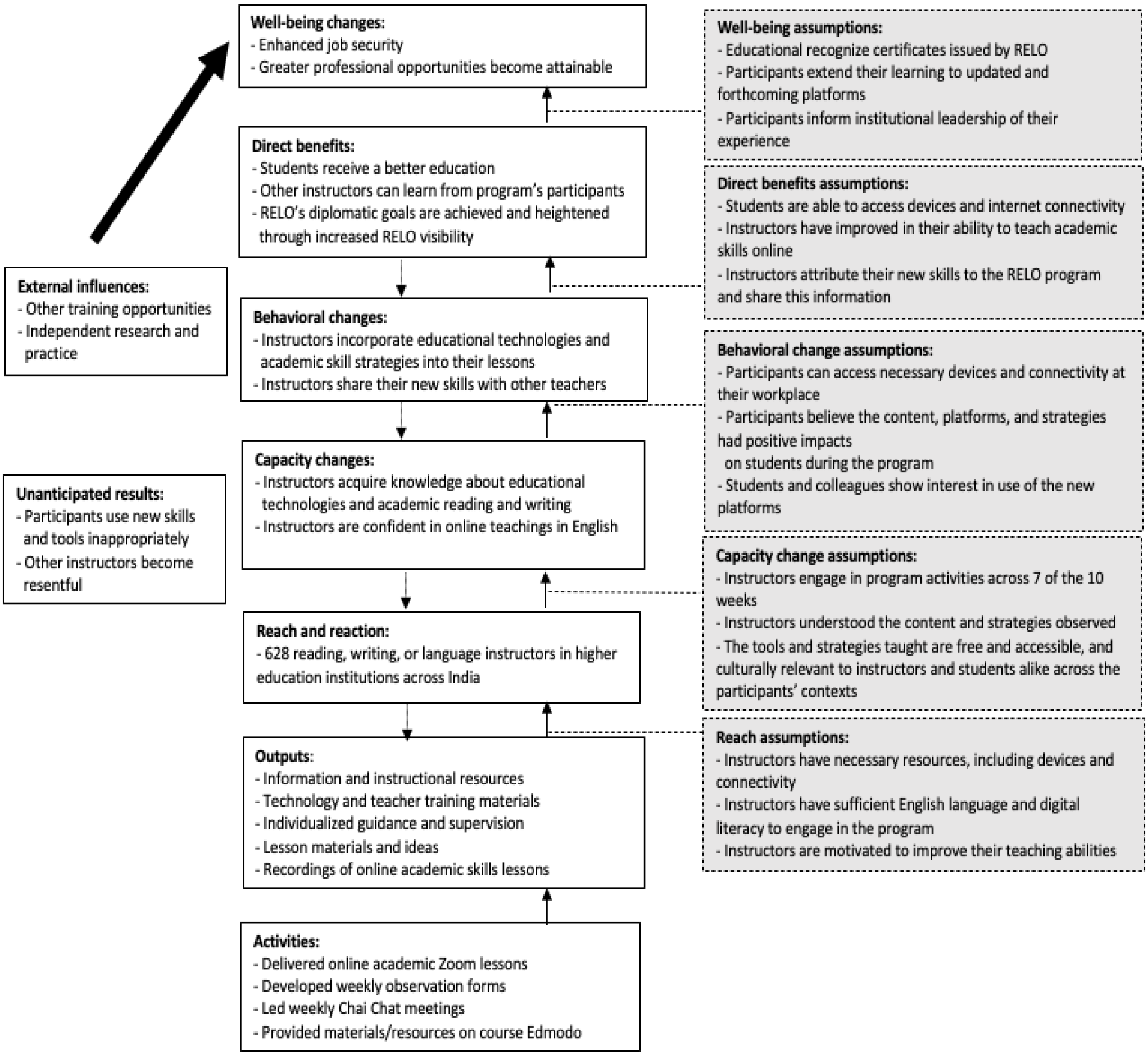
The attribution question here involves assessing whether impacts have been made in these areas, addressing whether any such developments are indeed the result of the intervening program, and, if they are, to what extent. The level of proof necessary is one of importance, as transitions to distance learning are imminent for most educators and likely to remain a reality to some extent for quite some time.

The intended contribution of the intervening program appears plausible. The problem is well-understood, though relevant baseline data exists only in the form of a pre-survey done by key stakeholders in assessing the direction that the intervening program should take, showing a need in HE PD regarding teaching with technology, along with opportunities to improve English skills and non-technology related instructional methods.

4.2. Step 2: Develop the postulated theory of change and risks to it

The ToC reflects all four anticipated outcomes throughout the results chain, alongside the assumptions underlying each link (Figure 1). Key stakeholders were involved in discussing these assumptions, signifying shared understandings and program goals among key players and the evaluator.

Figure 1. The Theory of Change



This ToC takes into account an intermediate level of detail while clearly demonstrating the expected contribution of the program. Strengths and weaknesses of the chain and its corresponding assumptions are discussed throughout the paper.

4.3. Step 3: Gather the existing evidence on the theory of change

4.3.1. Evidence on outputs and activities

The outputs of the intervening program have tangible evidence in numerous formats. Participants gained access to academic skills content-related material in the forms of activities on several free, web-based platforms—all copiable, editable, and shareable—with in-class and web-based tutorials to guide their use. They received links to other open-education resources (OERs), lesson plans, and exercises for each topic discussed. All videos remain accessible and shareable on RELO India's public Facebook website.

4.3.2. Evidence on assumptions

Existing evidence on the program's observed results, as well as ToC assumptions and the potential for external factors that may influence outcomes, are gathered and presented.

4.3.2.1. Reach assumptions

The program was expected to reach approximately 900 instructors across nearly all states of India. ECA affiliate offices advertised the program and recruited from HEIs within their respective regions. Instructors were informed that technological devices, stable internet connectivity, and bandwidth or data capabilities that support hours-long videos over ten weeks were required. This, unfortunately, likely excluded potential participants, particularly those living in underserved areas. 628 registrants were instructors in HE; as this paper aims to evaluate the program's value as a PD opportunity for tertiary instructors, only this data is included. Based on already-existing evidence gathered from the RELO office (Table 1), 128 of the original 628 (20%) registrants successfully completed the program, highlighting weakness in the *reach and reaction assumptions* link.

Table 1. Breakdown of completion status, registration, and survey respondents

Certificate of completion	Program registrants	Program evaluation respondents
Earned	128	96
Not earned	500	52
Total	628	148

In gauging the normalcy of such attrition, one may revert to the literature on similar programs. This is a free, voluntary program that results in no qualification, transferable credits, etc. Participants are not required to enroll nor complete the program, have not invested in it financially, and are intrinsically motivated. Though a completion certificate is issued, it is uncertain what, if any, advantage it carries in India's highly-competitive academic job market.

Massive Open Online Courses (MOOCs) seem similar in structure to this program: participation is typically voluntary; participant work is required, assessed, and receives feedback but courses are typically intended for skill development over certificates. Thus, literature on MOOCs may serve as a logical basis for comparison. Several hosts offer over 4,500 MOOCs courses (Bouzayane & Saad, 2017), whose high dropout rates are fairly well-documented. According to Li et al. (2016), Coursera, a highly popular MOOC platform, sees successful completion rates averaging 7% to 9%. Later studies reviewing other platforms found less drastic albeit high attrition rates for their MOOCs as well: Cobos et al. (2017) monitored attrition rates of two courses requiring a cumulative score of 60% marks to earn a certificate, and two others simply requiring completion of 50% of coursework. Average certificate issuance rates were 13.25% and 22.25%, respectively. Given that certificate requirements of the program evaluated here falls between these two in rigor—obliging 70% of attendance and task form submission that were reviewed and recorded but not scored—the expected completion rate would fall between them accordingly, as is the case with the 20.4% of HE registrants successfully completing the intervening program.

Also addressed in this link is the digital divide that undeniably exists within India. Additional statistics on Indian infrastructure per surveys from London-based company QS (PTI, 2020) show that only 15% of Indian households have broadband, 53% of whom report poor connectivity, with 11% reporting electricity issues. 40% of internet users depend on mobile hotspots, 96% of whom cited connectivity issues. Still, participants acknowledged accessibility requirements upon registration. Thus, reasons *why* this attrition occurred must be explored.

4.3.2.2. Capacity change assumptions

Participants' engagement throughout the program is already demonstrated through certificate issuance as outlined above. Statistics on certificate completion and task form submissions can be found in Table 1 and Figure 3, respectively. Access to all course resources remained free and accessible throughout the program and at present.

In assessing the likelihood that participants learned, understood, and adopt lesson materials demonstrated in the skills-based classes observed weekly, we first turn to the research. In their book on impacts of PD in education, Condon et al. (2016) describe how PD indeed benefits its participants in the long-term, which is shown to have positive trickle-down effects on student outcomes. However, when delving more deeply into the integration of educational technology, McDaniel and Kenny (2013) found that, though instructors often believe such tools support student learning, their instructional strategies do not as often reflect this—typically a result of their own lack of knowledge on its use. Therefore, this opportunity, with its guidance and demonstration of applied technology, is expected to be impactful.

4.3.2.3. Behavioral change assumptions

Participants' motivations seem intrinsic; their interest appears established, verifying this element of the link. Whereas participants' ability to utilize the educational technology in their current situation requires home accessibility to technology, their ability to extend use to traditional classroom settings requires institutional connectivity. Presently, data on accessibility specifically in Indian HEIs is not available.

The opportunity is positioned to be impactful based on its adherence to PD best practices. Research suggests that one-time PD events do not sufficiently provide participants with opportunities to effectively integrate technology into their own lessons (Gunter & Gunter, 2015; Lacey et al., 2014). Research from Gunter and Reeves (2017) found that PD opportunities that extend training across weeks and demonstrate authentic use of technology in teaching relevant content increase participant engagement; providing participants with opportunities to practice such strategies heightens their likelihood of adopting techniques. Such findings are corroborated by Darling-Hammond et al. (2017), who extensively reviewed 35 publications centered on positive links between PD opportunities in teaching and their effects on instructional practices and student outcomes. The authors suggest seven tips in best practices of PD programs: PD should be content-focused, deliver active learning opportunities, encourage collaboration, incorporate modeling of lessons, provide expert guidance, offer reflective feedback, and occur over an extended period of time. The intervening program involved all seven recommendations.

The attitudes of Indian educators toward educational web-based materials are important in gauging their likelihood of adopting them. In extensive efforts on behalf of UNESCO, Hodgkinson-Williams and Arinto (2017) review global perspectives—including Indian attitudes toward them—on OERs, which were heavily featured in the intervening program. Researchers found that, though Indian educators typically had little awareness of OERs, they were very receptive toward integrating them following engagement with them.

4.3.2.4. Direct benefits assumptions

Given adherence to best practices, it is likely that students will receive a better education, assuming they can access it. Though issues of Indian infrastructure have been addressed, a recent study from the University of Chicago has gathered accessibility data specifically relating to university students in India. Mukhopadhyay (2020) describes that 27% of university students from urban backgrounds have at-home internet access; this number drops as low as 2% for students from rural homes. This study suggests that 47% of university students have access to an appropriate device at home, though this number includes mobile devices—many of which are shared.

There are no publications regarding RELO's goals, nor of participants' experiences from previous Specialist programs. Yet, because of its importance to key stakeholders, it will be explored using new data.

4.3.2.5. Well-being change assumptions

Stated by Montague et al. (2002) as cited in Mayne (2008), links involving *indirect influence* are more difficult to validate, as is the case here. An exploration of the connection between PD, especially RELO certificates, and elevated job opportunities or security in India has not yet been undertaken. Nevertheless, Hodgkinson-Williams and Arinto (2017) found that educators in India do perceive a sense of heightened reputation after adopting and sharing OERs in their institutions.

4.4. Step 4: Assemble and assess the contribution story, and challenges to it

The contribution story is concisely detailed in Figure 2, as adapted from Mayne (2012), below:

Figure 2. Contribution analysis in evaluating existing evidence on program and stakeholder goals

Strengths of the assumptions chain:
- Developed a theory of changes with highly plausible assumptions
- Examined pre-survey data in informing intervention construction to confirm alignment of stakeholder goals and program delivery
- Reviewed program's activities and outputs in confirming the program's reliable, intended implementation and adherence to best practices in PD programming
- Established a research-grounded baseline for attrition rates of similar programs
- Detailed issues of infrastructure underlying India's digital divide within higher education
- Identified likely motivations for program participation
- Illustrated logical expectations for PD impacts in the Indian context
- Outlined research-based expectations for Indian attitudes toward applying PD-demonstrated practices

Weaknesses in need of further investigation:
- Explore explanations underlying high attrition rates
- Identify possible external influences
- Outline perceived relevance and benefits of the program
- Research the extent to which participants learned, understood, utilized program outputs
- Gather data on achievement of RELO goals

The ToC, though iteratively revised during program planning, is finalized as presented in step 2. The contribution story overall has both strengths and weaknesses, as reflected in Figure 2.

4.5. Step 5: Seek out additional evidence

This CA step centers around integrating evidence gathered in the weeks following program completion and aims to strengthen the ToC's validity. To assess whether the program contributed to changes in participants' teaching and achieving RELO's goals, further evidence was sought in the form of:
- participant program evaluations,
- student course evaluations,
- student post-surveys,
- Chai Chat Specialist questionnaires,
- weekly observation task forms, and
- interviews with key stakeholders.

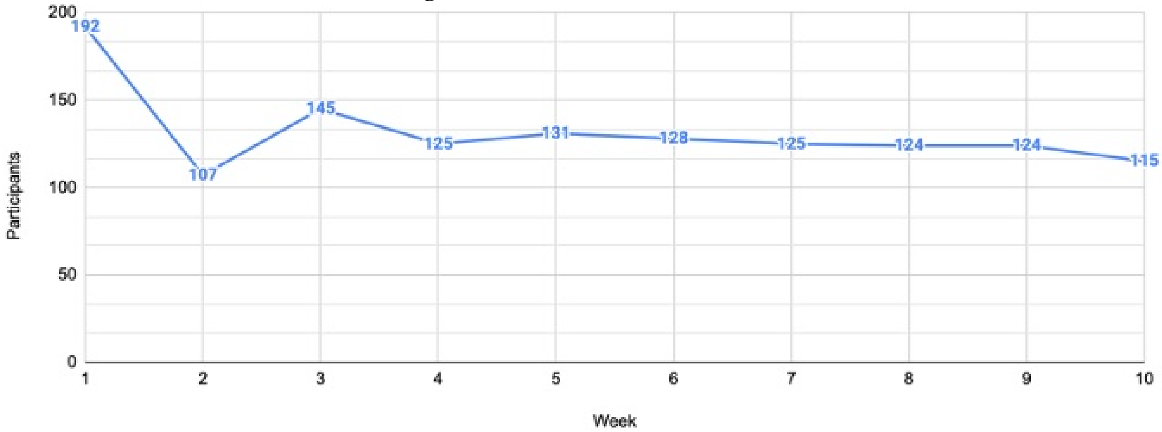
4.5.1. Reach and reaction

Evidence verifying this link involves deeper exploration of weekly observation task form submissions and program evaluation data.

4.5.1.1. Weekly observation task forms

As noted, the program faced high attrition rates. However, this is not uncommon for comparable initiatives, as outlined in step 3. Given these findings, the intervening program's attrition rates are fairly expected. Crucially, investigation of weekly task form submissions suggests the majority of participants dropped out before program initiation, based on comparison of registrants (628) to submissions of week-one forms (192). Figure 3 shows the number of observation task form submissions weekly, which average near the 128 issued certificates of completion (μ = 131.6).

Figure 3. Task form submissions



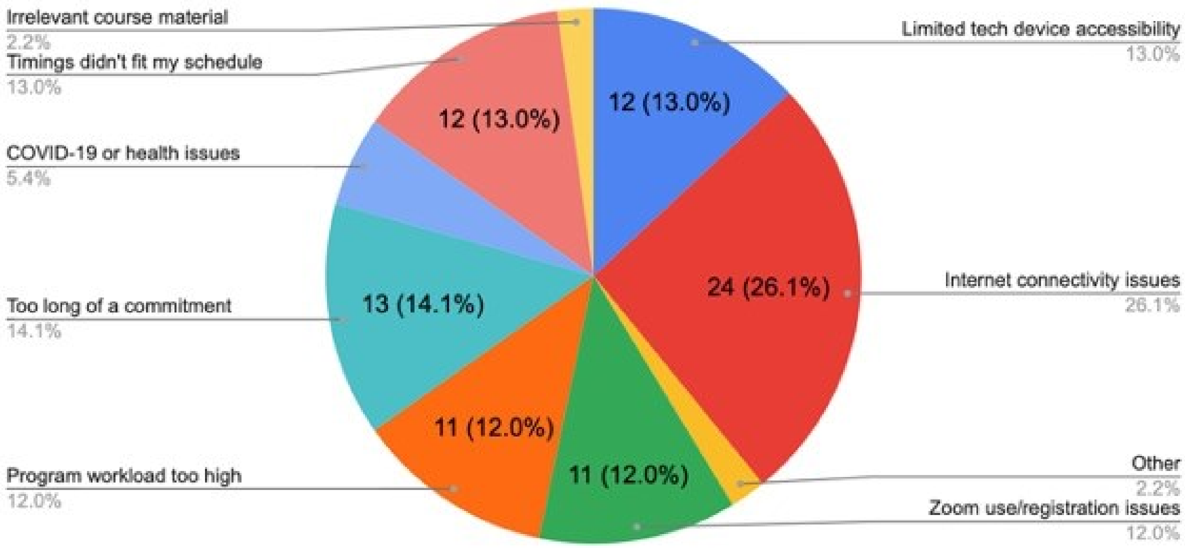
4.5.1.2. Program evaluation data of Non-CERT participants

Three weeks following completion of the Chai Chat program, participants were sent a program evaluation adapted from the SEEQ (Marsh, 1982). In total, 148 responses were received (n = 148) from educators in HE. Of these, 96 respondents had successfully completed the program (CERT group) and earned completion certificates. 52 respondents (Non-CERT group) did not receive a certificate (Table 1).

It is important to explore the reasons for attrition and to identify missteps as they correspond to assumptions in the results chain. Although the program evaluation was otherwise identical for all program registrants, the non-CERT group evaluation included an additional item inquiring about reasons for not completing the course.

Issues related to internet connectivity are cited (Figure 4) as the primary reason for attrition among the Non-CERT respondents (n = 52). Also commonly identified are issues of access to technological devices, and the digital literacy required to navigate registration to, access to, and materials within the course. Several responses address the timings of Chai Chat meetings (five per subject per week were offered) and the time commitment required of the program—valuable information for stakeholders in future planning initiatives. Only two individuals took issue with the content of the course itself.

Figure 4. Reasons for not completing the course (*Note.* Respondents selected all that apply)



Though this is strong indicative data, one cannot definitively state that the 52 Non-CERT respondents reflect the 500 registrants who did not earn a certificate nor the 436 who never began the program.

4.5.2. Capacity changes

The next link in the results chain addresses participants' growth in knowledge and ability, evaluated using program evaluation data of the CERT group. As non-CERT participants missed at least 40% of Chai Chat meetings, their data has been removed from calculations in the following sections; they likely gained insufficient exposure to course content.

4.5.2.1. Program evaluation data of CERT participants

The *capacity change* link appears to be well-verified by CERT participant data; continued participation may reflect their language and extended technology capabilities. According to responses on the program evaluation (Table 2), they have largely "learned and understood valuable… academic skills content, educational technology, and other materials" taught in the program. On a Likert-type scale of one (*Strongly Disagree*) to five (*Strongly Agree)*, Table 2 shows mean scores for each area of learning and percentages of respondents who *Agree* or *Strongly Agree.* Average scores typically reach about 4.0 (Marsh, 1982); means presented here and throughout the paper are often above 4.5.

Table 2. Participants' program learning

Instructional area	Mean score	% of responses of *Agree* or *Strongly Agree*
Academic skills content	4.65	96%
Educational technology	4.75	95%
Other resources	4.63	95%

4.5.3. Behavioral changes

The next link in the ToC's results chain investigates participants' changes to instructional practices. The program's impact on participants' teaching is evidenced again from program evaluations, along with post-surveys from students, student course evaluations, and questionnaires from Specialists, who led the Chai Chat discussions.

4.5.3.1. Program evaluation data of CERT participants

Extracted from program evaluations, data from the 96 respondents of the CERT group show participants largely self-report having incorporated both academic skills content and educational technology tools demonstrated in the Chai Chat program. Table 3 shows percentages of respondents who claim to have added program teachings into their lessons, and the percentage of respondents who have otherwise adapted teaching methods as a result of the program.

Table 3. Changes made to teaching as a result of the intervening program

Area of change	% of affirmative respondents
Added new academic skills content	92%
Added new activities with educational technology	93%
Adapted other elements of previous teaching strategies	95%

4.5.3.2. Student post-surveys

Students in the Indian HE system are often taught by the same professors throughout their education. All students participating in the skills-based courses were HE students whose lessons suddenly transitioned online at the previous academic year's end; they would therefore be capable of noting changes to instructors' online teaching and were familiar with content exemplified in the skills-based courses, uniquely allowing them to identify these methods and changes to their instructors' strategies.

Given that Chai Chat participants were invited from the same universities as students participating in the academic skills courses, these students were expected to have both instructors who did and who did not participate in the PD program and may then be able to compare approaches employed between them. However, of the 62 students who completed the survey, 50% were not sure whether their professors had participated or not.

Of the 31 remaining, 13 students (21%) representing six different institutions could definitively state having instructors who participated in the intervening program while 18 students (29%) reported that none of their professors were involved.

Of the 13 students who had participating instructors, nine (69%) stated having noticed differences to instructors' methods while the remaining four students (31%) reported no changes in instruction. Based on open-ended survey data, the changes seem to be in both content and educational technology; several students also commented that participating instructors led more interesting, motivating lessons and that the online environments they created included more interaction and engagement.

4.5.3.3. Student course evaluations

81 of the 160 HE students registering in the synchronous academic skills classes responded to program evaluations. Students' opinions of the course and its technology are pertinent in assessing whether the Chai Chat program provided relevant content for its participating instructors.

Table 4 shows, on a Likert-type scale of 1 (*Very Poor*) to 5 (*Very Good*), students' mean scores in the areas of content and technology and the percentage of respondents who highly assessed the value of the course. It also shows students' perceived abilities in academic reading and writing in English before and after the course and perceived self-improvement.

Table 4. Student perceptions of abilities and self-improvement

Program area	Mean score	% of *Good* or *Very Good* responses
Course Content	4.59	93.7%
Use of Technology	4.7	92.3%

Perceived ability before course	Perceived ability after course	Mean self-reported improvement	% of students reporting improvement
2.76	4.12	1.36	90.1%

4.5.3.4. Specialists' questionnaires

The program involved five Specialists, selected by the RELO office in and approved by the ECA, who led "Chai Chat" discussions with 15 to 30 individuals for one hour per week per academic skill.

Following the Chai Chat program's conclusion, Specialists were sent a brief questionnaire regarding their estimations of participants' integration of program information. Four of five Specialists responded (n = 4; 80%). All responding Specialists stated that participants had discussed using both the educational technology tools and the content presented in the skills courses. Specialists' approximations of how many participants incorporated program methods are detailed in Table 5.

Table 5. Specialists' estimations of participants who discussed program content, tools use

Specialist estimation, based on weekly	Specialist 1	Specialist 2	Specialist 3	Specialist 4
Chai Chat discussions	80%	60%	75%	70%
Average estimate of participants utilizing program material	71.25%			

4.5.4. Direct benefits

Here, it is proposed that students will receive a better education resulting from instructors' participation in the intervening program. Though difficult to verify, starting points for strengthening this link may include instructors' improvement in teaching academic skills, online, in English. A direct benefit for RELO and US government stakeholders involves the promotion of RELO programming, assuming participants attribute improvements to the program itself and are able and willing to share their new skills with colleagues.

Evidence that the program has positively impacted students' education has been gathered from CERT participants' program evaluations; to support goals of the RELO program, relevant data again from CERT participants' program evaluations are reviewed along with Specialists' questionnaire responses and data from interviews with two RELO representatives.

4.5.4.1. Program evaluation data of CERT participants

The assumption of improved education is rooted in the notion that instructors are providing better lessons. Participants were asked to self-assess their confidence in the areas of teaching academic skills content, online, and in English, both before and after the program. The majority of participants perceive improvement in all three areas, with details of mean perceptions on a 5-point Likert-type scale and percentages of respondents perceiving improvement in Table 6.

Table 6. Perceived improvement in teaching abilities resulting from intervening program

Instructional skill area	Mean confidence before the program	Mean confidence after the program	Mean increase	% of respondents perceiving improvement
Teaching academic skills	3.47	4.58	1.11	75%
Teaching online	3.35	4.57	1.22	75%
Teaching in English	4.12	4.60	0.48	57%

In addressing whether participants are able and willing to share information regarding RELO programs and opportunities, participants would have to consider the program valuable. When asked about their attitudes toward the PD opportunity overall, the mean rating for the program was 4.81, with 97% of respondents stating it was either *Good* or *Very Good*.

4.5.4.2. Specialist questionnaires

As the goals of the RELO office are most familiar to and best understood by those in close proximity to stakeholders, evidence of the program's value was gathered from Specialists and two RELO representatives involved in the project.

Whereas the four responding Specialists believe the program successfully advanced public diplomacy through promotion of the English language, only three believe the goals of mutual understanding through cultural exchange were achieved; the fourth reported being "unsure."

All four Specialists believe the program successfully advanced public diplomacy. Reasons listed for this include participants' praise of RELO programs, resources, and opportunities during Chai Chat meetings; gratefulness for opportunities to practice English; shared best practices in English instruction; and observations of a master English teacher.

Specialists are, however, divided on whether the program successfully fostered mutual understanding between the U.S. and India through cultural exchange. Those responding affirmatively cite discussion of shared experiences and challenges among teachers in both cultures, such as connectivity issues in rural areas and sharing one device for multiple residents within a U.S. household, to the surprise of the Indian instructors. Similarly, good teaching strategies, including those intended to overcome such challenges, are shared between these two groups and appear to be cross-cultural. One of the Specialists pointed out that the mutual understanding seemed to be occurring as a consequence of the program, as intended, but that it often seemed to emerge as an effect of "hits and misses." For example, many instructors of students in rural areas noted that they at times have to deliver content through televised or radio programs, in which the lead teacher had no experience or training; adapting lessons to fit India's most developing and underserved regions was left largely up to discussion during the Chai Chat meetings between the Indian participants themselves, though future programming could be modified in light of this.

One Specialist reported uncertainty on the program's success in this area, stating that students seemed to be responding more to cultural differences than instructors, and that several Chai Chat instructors "seemed completely lost," though they enjoyed the differences in approaches when grasped.

4.5.4.3. Interviews with RELO representatives

Two representatives of an India-based RELO-affiliated office, with over ten years' experience and greatly familiar with all regional RELO offerings, agreed to semi-structured interviews about the program's success. The purpose of the interviews is to discuss the fulfillment of RELO's goals of achieving public diplomacy through promotion of English and the fostering of mutual understanding between the U.S. and other countries through cultural exchange. Though Specialists were asked about their views, they have primarily worked in peripheral, limited capacities with RELO. The local on-site staff interviewed, conversely, have worked extensively and are perhaps more familiar with these programs in the Indian context than anyone involved in the project or evaluation. For their expertise and unique ability to compare this program's success to previous RELO India projects, their interviews are considered instrumental in validating this element of the *direct benefits* link.

Interviewee A is an Indian native directly involved in RELO's HE programs and is familiar with nearly all of its initiatives. She stated that she believes the intervening program did fulfill both RELO goals in that (A) it was clearly centered around the promotion of the English language, and (B) mutual understanding through cultural exchange had more of an opportunity to flourish, as it occurred to perhaps a greater extent with this program than similar projects, being the first RELO-India PD or teacher training program that spanned both time and country, and facilitated individualized communications between participants and Specialists regularly. She explained that, because of the asynchronous format of lesson observations, a larger number of professionals were able to engage closely with cultural components. The observation task forms often required participants to reflect on cultural aspects of the lessons, highlighting cultural exchange at a time when she believed it may have been otherwise overlooked within the lessons' quick-moving content. The small-group, social nature of Chai Chat meetings also allowed participants to discuss all aspects of a full course and cultural differences that emerged throughout.

Interviewee B is also native to India. She typically works with educators of economically disadvantaged students, providing support across secondary into tertiary education. Interviewee B called the response to the program "unprecedented," acknowledging that praise and participants of the program had spread into nearby countries such as Uzbekistan, Saudi Arabia, and Bangladesh, and that video links and resources had been shared as far as Peru. She believed this to be a clear demonstration of the success of public diplomacy through the promotion of the English language and a wide-reaching representation of American culture as reflected in education. Regarding mutual understanding, Interviewee B explained that, in order to assess the achievement of this goal, one must first illustrate the average English-language classroom in the Indian context as a basis for comparison, which typically consists of teacher-centered approaches, often heightened in distance learning when instructors have insufficient training in online pedagogies. The expert Specialists contracted for this initiative elicited the student engagement and interaction that has gained traction in other cultures in the realm of distance learning; participants were able to recognize the value in and observe real instances of applications of virtual strategies, making adoption of such techniques seem practical and possible, as opposed to idealistic, as is sometimes a concern following briefer PD programs.

4.5.5. Well-being changes

The assumptions underlying potential well-being changes are difficult to explore, as the intervening program under evaluation was completed so recently. In assessing whether participants are able to adapt to new or updated platforms, a certain amount of time must pass to allow for change to occur. It may perhaps be beneficial to follow-up with participants of the CERT group to investigate the topics of job security and/or new opportunities that stemmed from completing the program. In adding to the uncertainty of the validity of these assumptions, Interviewee A stated that the value of the RELO-issued certificates varies between organizations. To the best of her knowledge, they are most commonly used in fulfilling PD requirements of local institutions.

4.5.6. Evidence on other influencing factors

Potentially influential external factors include participants' involvement in outside PDs or trainings, or participants' self-guided research of and practice with new methods. To distinguish the impact of these factors from those of the intervening program, evidence is gathered from student post-surveys and program evaluations of the CERT group.

4.5.6.1. Student post-surveys

69% of student survey respondents who reported that their instructors participated in the intervening program stated they had noticeably changed their online teaching habits. All of the students who completed the survey (*n* = 62) also have instructors who did not participate in the program. Interestingly, only three of the respondents (12.5%) noted changes in the methods of professors that did not participate in the intervening program, possibly lending support to the idea that few instructors were motivated to independently research and incorporate new approaches into their lessons, though this indicative data is admittedly weak.

4.5.6.2. Program evaluation data of CERT participants

The program evaluation explicitly inquired about participants' involvement in other PD or training opportunities. 55 of 96 responding CERT participants (57%) reported engaging in other PD activities. Though this accounts for more than half of the participants, the topic of these opportunities was not addressed. Furthermore, questions in the evaluation largely emphasized growth "as a result of" the program and incorporation of tools and content "from this specific program." Similarly, when asked whether they felt more knowledgeable on web-based educational resources and teaching compared to colleagues who did NOT participate in the program, 92% of respondents answered "yes."

4.6. Step 6: Revise and strengthen the contribution story

In revising a strengthened contribution story, all steps are combined and presented succinctly in an effort to highlight the strengths and remaining weaknesses of the ToC links.

4.6.1. The revised contribution story

New evidence gathered since step 4 is presented to underscore this evaluation's areas of success and those needing additional exploration in Figure 5 below. As most links involve a mix of both weak and strong components, CA steps are instead categorized, as adapted from Mayne (2012).

Figure 5. Contribution analysis in evaluating program impact and achievement of governmental goals

Steps enacted via contribution analysis resulting in sufficient validation:

- Determined dropout timing and why dropout likely occurred
- Gauged participants' perceived learning, understanding, and improvements in and changes to teaching
- Confirmed program's usefulness and relevance via skills course student evaluations
- Gathered data from Specialists regarding achievement of program and RELO goals
- Conducted interviews with RELO employees regarding fulfillment of RELO goals
- Established an influence of external factors, including pressure from imminent transition to distance learning and outside PD opportunities

Steps enacted via contribution resulting in weak validation:

- Surveyed academic skills course students regarding professors' modified techniques
- Identified potential uses of completion certificates

Additional contribution analysis steps to consider in further validating the ToC:

- Survey participants' current or future students to assess sustained integration of program strategies, platforms, and/or content
- Inquire further on resulting professional opportunities and long-term benefits of completion certificates and the program overall
- Investigate possible effects on RELO visibility in the region, perhaps through analysis of social media traffic, numbers of registrants of upcoming projects, etc.

The revised contribution story has indeed been made stronger and more credible through the inclusion of program- and participant-specific data. Nonetheless, as pointed out in Mayne (2012), it is nearly impossible to develop a "foolproof" narrative.

4.6.2. Discussion of remaining weaknesses

The credibility of each link has been re-assessed for its evidence and logic. The program's *activities, outputs, reach and reaction,* and *capacity change* links are deemed strong and credible.

Though the *behavioral change* link remains fairly strong, there is relative weakness in evidence from sources beyond participants' self-assessments reported in the program evaluation. It is possible, also, that multiple students could be identifying changes in the same individual instructor(s) in the student post-surveys, which yielded relatively little evidence corroborating instructors' self-reports of instructional change in the first place.

The direct benefits link includes largely different intended outcomes from varying stakeholders, with the validation of some being more readily quantifiable than others so soon after program completion. Though self-perceived improvement in CERT participants' use of educational technology and instruction of academic skills provides somewhat strong evidence of improved education for students, more objective evidence in the form of quasi-experimental approaches would provide stronger support, if possible. Furthermore, whether heightened visibility of RELO India and its projects is verified remains uncertain.

As acknowledged in section 5.5, evidence for the *well-being changes* link is not yet available beyond Interviewee A's estimation that RELO certificates are recognized for individual institutions' PD requirements. Future data on job security and professional opportunities will likely be speculative, at best, in the form of reported perceptions in follow-up surveys and largely confounded with external influences over time. Essentially, this link is the weakest in the chain, with little to no evidence.

5. Discussion

This paper has presented a comprehensive evaluation of a multi-regional online teacher-training/PD Specialist project, funded by the US Department of State. The evaluation's purpose was to assess the project's impact in successfully influencing participants' online teaching practices in the areas of academic content and use of educational technologies, and in achieving RELO's goals of building public diplomacy through English programming and fostering mutual understanding between the U.S. and foreign countries through cultural exchange. The RUFDATA framework (Saunders, 2000) guided initial design phases of the evaluation in efforts to ensure full consideration across all program facets and key stakeholders, a number of whom were involved in the evaluation's development. Contribution Analysis (Mayne, 2001; Mayne, 2008) informed the data collection and analysis and was selected for its theory-based approach and applicability with programs that cannot feasibly be evaluated under experimental conditions. CA was iteratively employed in order to develop a plausible and validated Theory of Change (Weiss, 1995; Weiss, 1997) and to establish a strong, credible contribution story that served to infer causality of the program, as opposed to external factors, in adding value for stakeholders, as illustrated through a revised, comprehensive, and verified contribution story. This paper contributes to knowledge in that there are few if any formal evaluations of fully online teacher training or PD programs that provide participants the opportunity to both observe all aspects of an online course while also engaging with the instructor and other course contributors in small groups over an extended period of time, especially with populations straddling the digital divide. It is also the first peer-reviewed, published study, to the best of the author's knowledge, to formally evaluate a RELO-sponsored Specialist project despite the program's inception in 1991 and its current running rate of 150-200 international projects per year (U. S. Department of State, 2021).

Implications stem from both the findings of the evaluation and the evaluation's enactment. Teachers, teacher trainers, and professionals involved in planning similar programs—especially in developing regions and/or in purely virtual contexts—may note the successes and potential of distance training through observation and social connectedness while keeping in mind the impacts of challenges to infrastructure, even when disclosing accessibility requirements at registration, as well as registrants' digital literacy. Familiarity with techniques that key actors utilize on the ground to navigate such obstacles is key in maximizing participation and program value. High attrition is expected; eliminating caps on registrants or creating waitlists may help in reaching a greater number of educators.

The evaluation encountered a number of limitations. The evaluator overestimated the academic skills course students' knowledge of their instructors' participation in the intervening program, which led to weakness in the *behavioral changes* link. Evaluators should aim to further substantiate participants' self-reports in integrating program materials; follow-up with participants' current students can most accurately attest to lesson content, and comparison to feedback from instructors' previous course populations may yield truer findings. Furthermore, the

ToC proved to be complex and layered due to differing expectations of various stakeholders, including instructors and RELO representatives. This complexity may be better unpacked by applying an actor-based approach to CA as outlined in Koleros and Mayne (2019), in which researchers integrate multiple actor-based ToCs to develop clear, distinct images of actor-specific changes that should occur. Moreover, in continued iterations of CA, evaluators may investigate whether participants shared program resources or information regarding RELO initiatives with colleagues; assess RELO's potentially heightened visibility in studies on its social media traffic, reach, and future program registrations; ask for specifics on external PD activities; and inquire about professional opportunities resulting from participants' program experience.

References

Bouzayane, S., & Saad, I. (2017). Weekly predicting the at-risk MOOC learners using dominance based rough set approach. *European Conference on Massive Open Online Courses* (pp. 160–169). Springer.

Bureau of Educational and Cultural Affairs. (n.d.). *English Language Specialist program: At-a-glance*. Exchange programs. https://exchanges.state.gov/us/program/english-language-specialist-program

Careers360. (2020, March 21). *COVID-19: Students left in the cold as colleges shut down.* https://news.careers360.com/covid-19-students-left-in-cold-colleges-shut-down

Cobos, R., Wilde, W., & Zaluska, E. (2017). FutureLearn data: What we currently have, what we are learning and how it is demonstrating learning in MOOCs. *The 7th International Learning Analytics and Knowledge Conference* (pp. 1-20). Springer-Verlag Berlin Heidelberg. https://doi.org/10.1145/3027385.3029433

Coffey, M., & Gibbs, G. (2001). The evaluation of the Student Evaluation of Educational Quality questionnaire (SEEQ) in UK higher education. *Assessment & Evaluation in Higher Education, 26*(1), 89-93.

Condon, W., Iverson, E. R., Manduca, C. A., Rutz, C., & Willett, G. (2016). *Faculty development and student learning: Assessing the connections*. Indiana University Press.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute. https://learningpolicyinstitute.org/product/effective-teacher-professional-development-report

Gunter, G. A., & Gunter, R. E. (2015). *Teachers discovering computers: Integrating technology in a changing world* (8th ed.). Cengage Learning.

Gunter, G., & Reeves, J. L. (2017). Online professional development embedded with mobile learning: An Examination of teachers' attitudes, engagement and dispositions. *British Journal of Educational Technology, 48*(6), 1305-1317.

Hodgkinson-Williams, C. & Arinto, P. B. (2017). *Adoption and impact of OER in the Global South*. African Minds, International Development Research Centre & Research on Open Educational Resources. https://doi.org/10.5281/zenodo.1005330

Koleros, A., & Mayne, J. (2019). Using actor-based theories of change to conduct robust contribution analysis in complex settings. *The Canadian Journal of Program Evaluation, 33*(3), 292-315.

Lacey, C. H., Gunter, G., & Reeves, J. (2014). Mobile technology integration: Shared experiences from three initiatives. *Distance Learning, 11*(1), 1-8.

Lakshmi, G., & Agarwal, M. (2017). e-Readiness of state open universities: A Study of BRAOU and UOU. *Indian Journal of Open Learning, 26*(3), 177-202.

Li, W., Gao, M., Li, H., Xiong, Q., Wen, J., & Wu, Z. (2016). Dropout prediction in MOOCs using behavior features and multi-view semi-supervised learning. *2016 International Joint Conference on Neural Networks* (pp. 3130–3137). IEEE.

Marsh, H. W. (1982). SEEQ: A Reliable, valid, and useful instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology, 52*(1), 77-95.

Mayne, J. (2001). Addressing attribution through contribution analysis: Using performance measures sensibly. *Canadian Journal of Program Evaluation, 16*(1), 1-24.

Mayne, J. (2008). Contribution analysis: An Approach to exploring cause and effect. ILAC Brief No. 16: Rome: *The Institutional Learning and Change Initiative*. http://www.cgiarilac.org/files/publications/briefs/ILAC_Brief16_ Contribution_Analysis.pdf

Mayne, J. (2012). Contribution analysis: Coming of age? *Evaluation*, *18*(3), 270-280.

Mayne, J. (2015). Useful theory of change models. *The Canadian Journal of Program Evaluation, 30*(2), 119-142.

McDaniel, R., & Kenny, R. (2013). Evaluating the relationship between cognitive style and pre-service teachers' preconceived notions about adopting console video games for use in future classrooms. *International Journal of Game-Based Learning, 3*(2), 55–76.

Ministry of Rural Development. (2018). *Mission Antyodaya 2017-2018: Villages with availability of electricity for domestic use.* https://missionantyodaya.nic.in/ma2018/preloginStateElectricityReport2018.html

Ministry of Statistics & Programme Implementation. (2018). *Key indicators of household social consumption on education in India*. http://mospi.nic.in/sites/default/files/publication_reports/KI_Education_75th_Final.pdf

Mukhopadhyay, A. (2020). COVID-19: Online classes and the digital divide. *Ideas for India: For More Evidence-Based Policy.* TATA Centre for Development: University of Chicago. https://www.ideasforindia.in/topics/poverty-inequality/covid-19-online-classes-and-the-digital-divide-during-the-times-of-corona1.html

Naresh, B., Reddy, D., & Pricilda, U. (2016). A Study on the relationship between demographic factor and e-learning readiness among students in higher education. *Global Management Review*, *10*(4), 1-11.

Patton, M. W. (2008). *Utilization-focused evaluation* (4th ed.). Sage.

PTI. (2020, April 21). COVID-19: Indian Internet infrastructure not prepared for shift to online teaching-learning, says QS report. *Financial Times.* https://www.financialexpress.com/education-2/covid-19-indian-internet-infrastructure-not-prepared-for-shift-to-online-teaching-learning-says-qs-report/1935287/

Saunders, M. (2000). Beginning an evaluation with RUFDATA: Theorizing a practical approach to evaluation planning. *Evaluation, 6*(1), 7-21.

United Nations Educational, Scientific and Cultural Organization. (2020, June 18). *Education: From disruption to recovery.* https://en.unesco.org/covid19/educationresponse

U.S. Department of State (2021). *What is the English language specialist program?* ELPrograms. https://elprograms.org/specialist/

Wajnryb, R. (1992). *Classroom observation tasks: A Resource book for teachers and trainers.* Cambridge University Press.

Weiss, C. H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In J. Connell, A. C. Kubisch, L. B. Schorr & C. H. Weiss (Eds.), *New approaches to evaluating community initiatives: Concepts, methods, and contexts* (pp. 65-92). The Aspen Institute.

Weiss, C. H. (1997). Theory-based evaluation: Past, present, and future. *New Directions for Evaluation, 76*, 41–55.

Huang, B., Jong, M. S. Y., & Chai, C. S. (2022). The Design and Implementation of a Video-facilitated Transdisciplinary STEM Curriculum in the Context of COVID-19 Pandemic. *Educational Technology & Society*, *25* (1), 108-123.

The Design and Implementation of a Video-facilitated Transdisciplinary STEM Curriculum in the Context of COVID-19 Pandemic

Biyun Huang[1*], Morris Siu-Yung Jong[1,2] and Ching Sing Chai[2]

[1]Centre for Learning Sciences and Technologies, The Chinese University of Hong Kong, Hong Kong SAR, China // [2]Department of Curriculum and Instruction, The Chinese University of Hong Kong, Hong Kong SAR, China// lucyhuang99@cuhk.edu.hk // mjong@cuhk.edu.hk // cschai@cuhk.edu.hk

[*]Corresponding author

ABSTRACT: The COVID-19 pandemic has brought disruptions and constraints to K-12 STEM education, such as the shortened classroom time and the restrictions on classroom interactions. More empirical evidence is needed to inform educators and practitioners which strategies work and which do not in the pandemic context. In response to the call for more empirical evidence and the need for cultivating responsible and competent 21st century citizens, we designed and implemented a transdisciplinary STEM curriculum during the COVID-19 outbreak. In order to facilitate the smooth delivery of the learning contents and authentically engage learners in the learning process, multi-model video approaches were employed considering the characteristics of three disciplines, STEM, social service, and writing, as well as learner diversity. Pre- and post-test results indicated that students' transdisciplinary STEM knowledge improved significantly after completing the curriculum. The integration of STEM, social service, and writing disciplines promoted the growth of students' empathy, interest, and self-efficacy. Consistent with the quantitative results, students responded in the interview that their STEM knowledge and empathy were both enhanced. Some implementation strategies introduced in the current study are also applicable when the standard teaching order is restored in the post-COVID-19 era.

Keywords: STEM, Transdisciplinary STEM, Video-facilitated approach, Social service, Empathy

1. Introduction

Due to the COVID-19 outbreak, over the world, students are unable to attend schools as per the previous norm. Consequently, emergency education is put into practice in many countries (Bozkurt et al., 2020). To reduce the loss of curriculum time, the Hong Kong Education Bureau has requested all subjects to make a series of adjustments to standard scheduling procedures. In the context, the implementation of transdisciplinary STEM education faces many challenges, such as the reduction of course capacity due to the shortened classroom time, and limited classroom interactions for maintaining social distance. In response to the emergency, many STEM disciplines moved to online learning and used video-based learning approaches to ensure content delivery. For example, in the United States, the urology residents training was changed from didactic sessions to video-based online sessions (Tabakin et al., 2021). In other universities, instructors used pre-class video sections to prepare Chemistry students for subsequent synchronous Zoom lectures (e.g., Lapitan et al., 2021). However, empirical studies examining the effectiveness of the video-facilitated instructional approach in the pandemic chiefly centered on the higher education sector. At the K-12 level, there is a clear emphasis on knowing how to organize STEM courses smoothly in the COVID-19 context. It is essential to know how STEM was carried out during the pandemic, what impacts it achieved, and what lessons can be learned.

Currently, at the K-12 level, there are several studies examining student and parent perceptions on distance learning regarding the adequacy of online learning materials (e.g., Chang et al., 2020; Fiş Erümit, 2020), the collaboration styles during homeschooling (e.g., Yates et al., 2020), and teacher perspectives on technology-enabled remote learning (e.g., Ewing & Cooper, 2021; Jong, 2019a). These studies enabled us to understand, at the macro level, what worked in emergent distance education and what needs to be improved. Nevertheless, in terms of specific K-12 disciplines, the collected empirical evidence is insufficient. Not to mention transdisciplinary STEM, which needs the collective efforts from multiple domain experts.

Transdisciplinary STEM refers to the production of new perspectives and solutions to problems by drawing upon multi-discipline knowledge and skills (Gibbs, 2015). Many problems in the natural world are complex and could not be solved with knowledge from a single discipline. Hence, drawing on the expertise of multiple disciplines can assist in developing a more comprehensive understanding of the situations and create new possibilities for solutions (Quigley et al., 2019). In K-12 STEM education, it is widely agreed that empathy, care, and STEM education should be integrated to cultivate 21st-century citizens who would develop socially responsible and environmentally sustainable solutions (Rulifson & Bielefeldt, 2017; Gunckel & Tolbert, 2018). Lee and Campbell (2020) proposed an instructional framework advocating the use of science and computer science content related to COVID-19 to engage K-12 students in understanding the phenomena and solving societal problems. Several studies made efforts to integrate STEM education with the element of empathy (e.g., Hutchison, 2016). However, to the best of available knowledge, none of them examined the effectiveness of using a video-facilitated approach to ensure the continuity of transdisciplinary STEM education in the COVID-19 context.

This study aimed to design a video-facilitated transdisciplinary STEM curriculum and test its effectiveness in a secondary school. It is intended as an empirical reference to the implementation of transdisciplinary STEM during the pandemic. In addition, the video-facilitated approach in this study is generalizable to other STEM courses when standard teaching order is restored in the post-COVID-19 era because the difficulties discussed in this paper (e.g., catering to individual learner differences) exist in both normal and non-normal learning settings in STEM education (Epler-Ruths et al., 2020; Jong, 2019a; Jong et al., 2020).

The research questions are:

- In the COVID-19 context, what is the impact of the video-facilitated transdisciplinary STEM curriculum on students' factual knowledge?
- In the COVID-19 context, what is the impact of the video-facilitated transdisciplinary STEM curriculum on students' design competence?
- In the COVID-19 context, what is the impact of the video-facilitated transdisciplinary STEM curriculum on students' empathy, self-efficacy, and interest?
- In the COVID-19 context, what are students' perceptions of the video-facilitated transdisciplinary STEM curriculum?

2. Literature review

2.1. Transdisciplinary STEM learning

Transdisciplinary STEM applies knowledge and skills learned from two or more disciplines to real-world problems and contributes to an active learning experience (Vasquez et al., 2013; English, 2016). There has been a rising call to emphasize connections between disciplines in STEM education (Chai et al., 2020; Geng et al., 2019; Honey et al., 2014; So et al., 2020). For example, in the United States, the STEM Task Force Report (2014) emphasized that STEM education is not a convenient integration of four disciplines, but the incorporation of real-world problem-based learning that connects the disciplines through coherent and active teaching and learning practices. Many scholars believe that the best preparation for students' future careers must involve interdisciplinary thinking (Quigley et al., 2019). Having STEM taught in a more connected way and set in the context of real-world problems will make STEM subjects more beneficial to students. This practice could lead to increased motivation, improved achievements, and higher persistence (Honey et al., 2014). In turn, these outcomes will help meet the call for a robust workforce (Linnenbrink-Garcia et al., 2016).

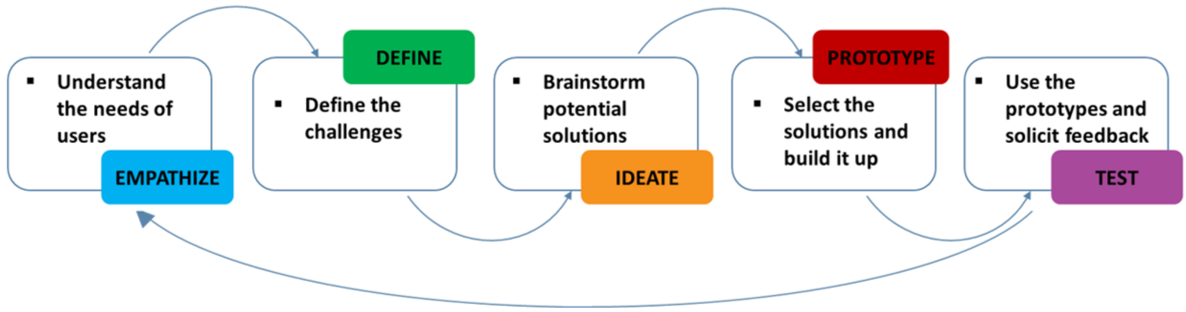
However, it is more common to integrate two or more science-related disciplines in STEM education rather than integrating science with disciplines from social science. For example, Hong et al. (2019) integrated scientific inquiry, mathematical thinking, and design technology into a college-level STEM course and promoted learning through a knowledge-building forum. They found that students demonstrated their competence in designing and improving designs in the knowledge-building environment. The study contributed to our knowledge of a practical approach for enhancing STEM learning, but it did not cover the social science discipline. Maiorca et al. (2021) studied the impacts of an interdisciplinary summer school project. The project combines science, education, medicine, and engineering to give Grade 5 to Grade 8 students authentic and hands-on STEM learning experiences. They found that this program enhanced students' self-efficacy and interests. While there are many socio-scientific issues, such as various forms of pollution and fighting the pandemic, which requires an integrated understanding of STEM and humanities, few studies integrate the two fields (Gunckel & Tolbert, 2018).

2.2. Theoretical framework for STEM curriculum design

To foster connections between STEM and humanities, the design thinking framework proposed by the Hasso Plattner Institute of Design at Stanford University is a reasonable choice. Beginning with empathy, the framework naturally draws on disciplines such as social studies to identify problems that confront humanity. Empathy is also a psychological construct that could motivate students to learn engineering knowledge (e.g., Chai et al., 2020). Grounded in emphatic understanding, designers then define the problems, ideate, prototype, and test (Hasso Plattner Institute of Design [HPID], 2010). For simplicity, we refer to the design thinking model as EDIPT. The EDIPT model is widely accepted in the design and STEM education fields. It advocates the necessity to understand users' potential needs with an empathic mind, and then work out solutions to cater to the needs. When needed, subsequent improvement of the prototypes is processed based on feedback from the users. Liedtka (2018) commented that the organized design process of the EDIPT model could help innovators carry out design processes in a more systematic manner and provide them with a sense of psychological safety to experiment.

As depicted in Figure 1, there are different focuses in the five phases of the EDIPT model. In the empathy phase, students visit users and talk with them to understand their potential needs. In the definition phase, students can synthesize and select the needs they consider important to fulfill, and then identify the one they will focus on in their design. In the ideation phase, learners propose a range of possible solutions to choose from by applying divergent thinking. Students build a prototype of the solution to bring them closer to their final solution in the prototype phase. In the testing phase, students demonstrate the prototype to users to collect feedback and further refine the solution. Simeon et al. (2020) applied the EDIPT model in a secondary school to promote the learning of physics concepts and found that both female and male students improved their achievements in physics at the completion of the course. Morrin and Liston (2020) implemented the EDIPT model in a STEAM project where arts and design thinking were promoted involving pre-service and in-service teachers and elementary school students. The project generated positive impacts on the attitudes and competencies of the teachers and students. These outcomes demonstrate that the EDIPT model is appropriate for scaffolding the design of secondary school STEM courses. Echoing the cross-disciplinary calling for the combination of technology and human-centeredness, we designed an integrated curriculum that builds on three subjects, STEM, social service, and writing, the latter of which has been less studied in the literature.

Figure 1. EDIPT Model and key procedures adapted from HDIP (2010)



2.3. Video-facilitated learning approach

Educational videos have become important content delivery tools for K-12 and higher education with their widespread application in flipped, blended and online courses (Brame, 2016; Jong, 2019b; Lin & Chen, 2019). Video lectures often allow students to fully comprehend the course material by allowing them to playback the video content as often as they need to, thus catering to students' individual differences (Brecht & Ogilby, 2008; Chen & Wu, 2015; Song et al., 2017). The advantages of video are substantial, such as (1) demonstrating the procedure for using a tool or equipment; (2) presenting the dynamics of a change or principle of motion; (3) replacing field trips with precise visual images of a scene and giving or providing students with a sense of immersion or facilitating a sense of student immersion; and (4) increasing the interest of the course by connecting it to real-world problems (Bates, 2019). Furthermore, a subset of video-based learning, the flipped learning approach, is beneficial to learning in a number of ways (Bond, 2020). The pre-class videos could free up in-class time for learning (Lo & Hew, 2021), potentially reduce the perceived course difficulty by introducing relevant concepts before class (Bond, 2020), and empower students to take ownership of their learning (D'addato & Miller, 2016). To ensure the effectiveness of video designs, Brame (2016) recommended three principles to follow: managing cognitive load, maximizing student engagement, and promoting active learning. According to Brame (2016), adding signal words or colors to highlight important information, keeping the video brief, using conversational language, and using guiding questions are effective strategies in engaging students.

In STEM subjects, video-based learning approach is often adopted to facilitate learning. Lo and Hew (2021) applied video-based flipped learning in middle school mathematics courses and found that students achieved significantly higher learning gains than the class without pre-class videos. Students in the video-based learning group also reported that the pre-class video boosted their confidence in the in-class problem-solving session. Similarly, Tsai et al. (2020) used pre-class videos in a middle school civic education class and reported that the videos promoted students' performance and learning motivation. Jong et al. (2020) compared the video-based virtual reality approach with the traditional textbook-based approach, and found that the video-based virtual visits to natural environments enabled students to connect knowledge to authentic contexts and achieve better learning outcomes.

In the context of COVID-19 and the resultant closure of many schools, countries and regions are using educational videos as one of the primary tools for content delivery (Pal & Patra, 2020). In Algeria, for example, the ministry of education has launched a YouTube channel and uploaded curriculum-related videos for K-12 students to study at home (Bozkurt et al., 2020). In the UK, Conlon and McIntosh (2020) found that student nurses perceived videos demonstrating scenarios held more authenticity and social relevance than digital audio and photobook styles. In Malaysia, pre-recorded lectures and hands-on training sessions were used for medical physics education during the partial lockdown. Students reported that short videos with questions helped them understand the topics better than the lengthy ones (Azlan et al., 2020).

Building on the experiences shared by scholars and researchers, we designed the video-facilitated instructions for students of the STEM curriculum. Since this is a transdisciplinary course and each course has unique characteristics and roles, we designed the video strategy based on each course's characteristics. The design will be described in the methods section.

3. Methods

3.1. Design of the video-facilitated transdisciplinary STEM curriculum

The transdisciplinary curriculum design is informed by the notion that integrating STEM and social studies provides an authentic context for problem-solving. Authentic problems such as the difficulties faced by visually impaired people in daily life were presented to students. To resolve these problems, students need to draw on their existing knowledge from multiple disciplines, such as user needs, materials, product design, tools and platforms. The design followed the process in EDIPT model (HPID, 2010). The prototype and test stages were delayed with a written proposal and presentation, where students presented their ideas verbally and collected feedback from experts and peers. The progression from "empathize" to "feedback" should help students understand the basics of designing a user-centered solution. In writing the proposal, students learn the language knowledge and skills for presenting the design solution proposed in the social service course. Additional content covered in the STEM course includes the fundamentals of electronic circuits, coding through Blynk and Thunkable, and the Internet of Things (IoT). Table 1 illustrates the main contents covered in the curriculum. Figure 2 depicts the in-class STEM activities.

Table 1. Contents of the transdisciplinary STEM curriculum

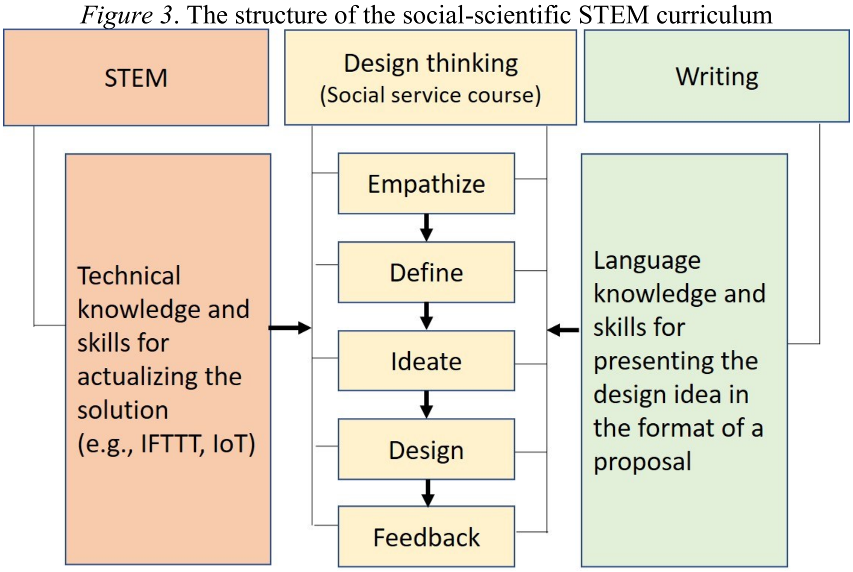
Discipline	Duration	Main purpose	Main contents/ Topics
STEM	6 weeks (1 hour/lesson)	Technical knowledge and skills	(1) Basic coding skills and computational thinking. (2) Basic IoT concept and applications (1). (3) Basic IoT concept and applications (2). (4) IFTTT and Smart Home Device. (5) Project-based Learning: Maker Education. (6) Mini Project on STEM Education.
Social service	5 weeks (1 hour/lesson)	Design thinking process & background knowledge	(1) Basic understanding of social services. (2) Basic knowledge of the user group and services provided by social service organizations. (3) Basic concepts and skills for developing a product for the user group. (4) Video-based field visit and (deeper) understanding of user needs. (5) Presentation of initial solutions and feedback from experts and peers.
Proposal writing	3 weeks (30 mins/lesson)	Language knowledge and skill	(1) Introduction of proposal content and wording. (2) Writing a proposal for the sample product. (3) Writing a proposal for students' own design solutions.

Figure 2. In-class STEM activities

STEM teacher teaches class through live broadcast | Students work on the hands-on sessions in-class | Sample student coding work on Thunkable

The curriculum implementation consists of two stages. The first stage was described above. The second stage involves prototyping and testing the solution. Due to the pandemic, students needed to maintain social distance, and could not work close to prototype products. The prototype creation and testing would be carried out when social-distance restrictions are eased. Hence, we are reporting the implementation and outcomes of the first cycle. See Figure 3 for the design of the first stage transdisciplinary curriculum.

Figure 3. The structure of the social-scientific STEM curriculum



In terms of the organization of teaching activities, we considered several issues, such as

- The shortening of teaching time since only half-day classes can be conducted.
- The social distancing and the unavailability of group work in class make learning more challenging for individual students.
- The student diversity (e.g., different speeds in coding) should be addressed.

In response to these challenges, a series of alternative video-facilitated strategies have been adopted. The level of student engagement with videos differs across course styles (Guo et al., 2014). Hence, we tailored the video strategies for each course accordingly. In the STEM course, short videos of less than three minutes were provided for students to preview before class. Introducing concepts before lessons can increase the active learning time in class (Lo & Hew, 2021). The length of the pre-class video was purposely shortened to make it more engaging (Azlan et al., 2020). For the hands-on sessions, all procedures were pre-recorded so that students could watch the video while working to offset the problem of not maintaining pace with the instructor during the class. Students can pause and revisit the video in their own time as required (Yates et al., 2020), and reduce their perceived course difficulty (Bond, 2020). Upon completing a hands-on section, students would take a picture of the completed work and upload it to Google classroom. Then, in the social service classroom, students watched the field visit video and completed the questions on the worksheet. Though virtual field visits may not be as informative as actual field visits, they can add variety to the learning experience and enhance authenticity (Chang et al., 2020; Friess et al., 2016). The video-based field visits were intended to make the topic more approachable and help students gain a deeper understanding of the context (Conlon & McIntosh, 2020; Jong et al., 2020). See Figure 4 for the typical video approaches applied in the study. Table 2 introduces the educational purposes of the approaches. Details of the video-facilitated strategies in the COVID-19 context are illustrated in Appendix I.

Figure 4. Screenshots of typical video-assisted approaches adopted in the study

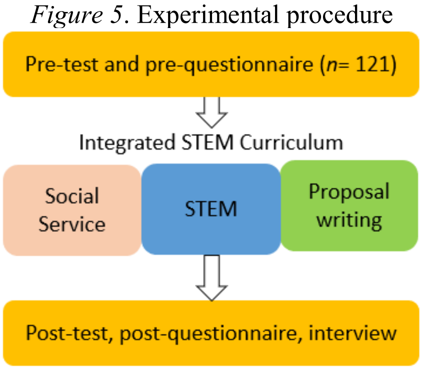
Table 2. The video approaches and educational purposes

	Video approaches and educational purposes
STEM	(1) Pre-class short videos: To introduce the basic concepts of each topic. (2) In-class video of hands-on sessions: To attend to learner differences and make sure learners with different learning paces can keep up with operation progress. *Note.* The length of each video was less than 3 minutes. Videos longer than 3 minutes were split into several clips.
Social service	(1) In-class video of field visit: To provide students with a sense of authenticity and make connections between learning and real-world problems. To provide students access to gain a close understanding of the needs of users (e.g., the visually impaired group) (2) Post-class video of extension videos. To provide students with access to know more assistive technology tools. *Note.* The in-class video was filmed before class by the project team. It presented the interview between the project team, several visually impaired people, and the social workers in the social service organization. Guiding questions on worksheets were assigned to students in class.
Proposal writing	(1) In-class video of an exemplar product's development background: To provide a sense of authenticity and link learning with real-world examples. To introduce the back story of an assistive technology tool. (2) In-class video of product demonstration: To introduce the aspects to be included in a product proposal. *Note.* Guiding questions were assigned to students in class.

3.2. Participants

To examine the effectiveness of the curriculum, we conducted a single group pre- and post-test experiment. It tracked the changes in students' transdisciplinary STEM knowledge, empathy, self-efficacy, and interest after completing the course. In addition, interview results with students were analyzed to triangulate the data. In total, 121 students gave consent to participate in the research. Their ages were between 12 and 14. The participants were from four classes of Grade 8 in a secondary school. Among them, 49 were females, and 72 were males. As illustrated in Appendix I, one STEM teacher taught four classes simultaneously via live streaming. At the same time, each class was accompanied by an experienced teacher and a student mentor to support students on-site. All the learning materials, such as hands-on practice videos and e-handouts (i.e., steps of the hands-on session and the flow of the course), were released to students before class on Google classroom. The experimental procedure is presented in Figure 5.

Figure 5. Experimental procedure



3.3. Measuring tools

3.3.1. Pre- and post-test of knowledge

3.3.1.1. Pre- and post-test of factual knowledge

The knowledge test was self-constructed based on the content covered in the integrated curriculum. The test comprises a series of factual knowledge questions about STEM and social service. The pre-test consisted of seven questions testing factual knowledge and one design challenge. For example, in testing STEM factual knowledge, one of the questions was:

> Which of the following is an input device?
> A. Buzzer, B. LED Screen, C. Infrared sensor, and D. Motor.

In testing social service knowledge, one of the questions was:

> What technology products do you know of that can help the underprivileged people in the community (e.g., people who are visually impaired or have poor living conditions)? Please list one product.

To ensure the validity of the test, the STEM part and the social service part of the test were drafted by one experienced STEM teacher and one experienced social service teacher of the project, and then reviewed by two experts in the field (Moss, 1992). After revision, it was further reviewed by two secondary school teachers to ensure readability. To establish internal reliability of the test scores, the test was marked by two scorers following a marking scheme with examples. The first rater trial scored 20 submissions, and then scored all the remaining submissions. The second rater randomly selected 30% of the papers for scoring, and the scores were compared with the first rater. The percent agreement (Campbell et al., 2013) between the two scorers was 92%. The same set of questions was used for the post-test to be consistent in understanding students' changes after completing the curriculum. It ought to be noted that the pre-test was done in the classroom, and the post-test was organized through Zoom due to the outbreak of another wave of disease in Hong Kong. Students were informed that the test results impact their course grades.

3.3.1.2. Pre- and post-test of design competence

In the design challenge, students were given a situation and were asked to propose a solution to the problem. This approach was inspired by Atman (2007), who used the challenge of designing a playground for the neighborhood to compare the design competence of undergraduates and expert designers. In the study, as the learning content was focused on social service and STEM, we evaluated students' design competence in designing a traffic light for the visually impaired group. The description of the design challenge is presented below:

> If you are a product designer and need to design a traffic light for the visually impaired, how would you design this traffic light? Please list: (1) the features of the product; (2) the main functions; (3) the input and output devices that will be used; (4) the reasons behind this design; (5) and introduce the design with a picture.

The marking scheme refers to the scoring method of scientific problem solving proposed by the Organization for Economic Cooperation and Development [OECD] (2019). The OECD (2019) uses a three-level system for evaluating students' scientific problem solving, i.e., full points for appropriate and original, partial points for appropriate only, and no points for all other cases. The advantage of using the system is that this criterion is easy to understand by the scorers with minimum training, and thus is more likely to promote the reliability of the scoring results. The same design challenge was used in the post-test. The marking scheme for design competence is shown in Table 3.

A marking guideline with examples was presented and introduced to the first scorer. After a trial scoring of 20 submissions, the first scorer marked all the remaining submissions. Then, another scorer randomly selected and marked 30% of the submissions. The percent agreement (Campbell et al., 2013) between the two scorers was 90%.

Table 3. The marking scheme for design competence

Dimensions	Corresponding sub-questions	Scores
Defining design goal	Features & Main functions	• 2 points: Correct, reasonable • 1 point: Partially correct, partially reasonable • 0 point: Unreasonable, incorrect
Technical knowledge/skill(s)	(1) Input device	• 2 points: Correct, reasonable • 1 point: Partially correct, partially reasonable • 0 point: Unreasonable, incorrect
	(2) Output device	• 2 points: Correct, reasonable • 1 point: Partially correct, partially reasonable • 0 point: Unreasonable, incorrect
Reasoning	Explanation of the design rationale	• 2 points: Correct, reasonable • 1 point: Partially correct, partially reasonable • 0 point: Unreasonable, incorrect
Visual presentation	Picture of the design	• 2 points: Correct, reasonable • 1 point: Partially correct, partially reasonable • 0 point: Unreasonable, incorrect
Creativity	Overall design	• 2 points: Reasonable and different from the solutions of most students • 1 point: a. Partially reasonable, different from the solutions of most students; or b. Partially reasonable, proposed two or more solutions but similar to other students' solutions • 0 point: Similar to the solutions of most students.

3.3.2. Survey questionnaire

The questionnaire consists of 12 items in 3 dimensions, including empathy, self-efficacy, and interest. The questionnaire employed a 6-point Likert scale ranging from "1- strongly disagree" to "6- strongly agree". In the dimension of empathy, the items were adapted from the instrument of Vossen et al. (2015), which measures people's empathetic mindset and sympathy. One sample item of the empathy dimension is "When people talk about how they feel about community service, I listen attentively." In the dimension of self-efficacy, the items were adapted from the instrument of Chen et al. (2001), which examines people' beliefs in their capabilities. A sample item of the self-efficacy dimension is "I believe I can design a good STEM solution to improve community service." In the dimension of interest, the items were adapted from the instrument of Luo et al. (2019), which evaluates people's interest in different subjects. A sample item of the interest dimension is "I want to learn as much STEM knowledge as possible." To understand the changes of students in emotion and motivation after completing the course, we pre- and post-questionnaire. The newly assembled questionnaire was subjected to expert review (Moss, 1992) by three university professors for face validity. After revision, it was further reviewed by three secondary school teachers to ensure readability.

3.3.3. Student interviews

To gain a deeper understanding of students' perceptions of the transdisciplinary curriculum, we invited 14 students to participate in 4 group interviews with their mother language. Each interview involved 3 to 4 students, lasted for 30-50 minutes. All interviews were audio-recorded and transcribed. The interview questions are: "(1) What is your overall feeling about the curriculum?; (2) Which part of the curriculum do you like best? Why?; (3) Which part of the curriculum do you think needs improvement? Why?; (4) Do you expect to receive any extra support?"

4. Results and discussions

4.1. Pre- and post-test of knowledge

4.1.1. Analysis of factual knowledge

To maintain the consistency of data analysis, only the students who participated in both the pre- and post-tests were included for data analysis. In total, 83 students participated in both the pre-test and post-test. In terms of factual knowledge scores, a paired sample *t*-test was conducted to examine if there was any difference between students' factual knowledge scores before and after the project. The results indicated that there was a significant difference in the factual knowledge scores for pre-test ($M = 8.41$, $SD = 3.45$) and post-test ($M = 9.63$, $SD = 3.51$); $t(82) = -2.64$, $p = .01$. See Table 4.)

Table 4. Pre- and post-test results of factual knowledge and design competence

		Mean (*SD*)	*n*	*t*	*p*-value
Factual knowledge	Pre-test	8.14 (3.45)	83	-2.64	.01
	Post-test	9.63 (3.51)			
Design competence	Pre-test	5.04 (1.94)	44	-3.10	< .001
	Post-test	6.33 (2.71)			

4.1.2. Analysis of design competence

Similarly, only students who submitted both the pre- and post-designs were included for data analysis. In total, 45 students submitted both the pre- and post-designs. A paired sample *t*-test was conducted to examine if there were any differences in students' design thinking scores. The results indicated that there was a significant difference in the design thinking scores for pre-test ($M = 5.04$, $SD = 1.94$) and post-test ($M = 6.33$, $SD= 2.71$); t $(44) = -3.10$, $p < .001$ (see Table 4). As mentioned in the research methods part, the post-test was organized via Zoom because of the outbreak of another wave of disease. Some students may not have uploaded their design pictures due to the inconvenience of doing so, and some students may have skipped this part due to the complexity of the task. This is a limitation of the study.

The test results demonstrated student improvement in both factual knowledge and design competence. The results are more positive than several preceding studies in the COVID-19 context, which reported a loss of learning amongst their findings (e.g., Engzell et al., 2021). There are two possible reasons for the improved outcomes. One is that the EDIPT model gave students ample opportunities for inquiry. It enabled them to integrate all the lessons they learned with solving social problems. Thus, it triggered students' interest in STEM and design. In this process, students build up their knowledge step by step. The results are consistent with the previous finding that connecting learning with real-world problems led to increased interest and achievement (Linnenbrink-Garcia et al., 2016). A second possibility is that it was facilitated by adequate and timely support. For instance, in the STEM course, videos for the hands-on sessions were made available, so if students could not keep up with teachers' pace, they could follow the videos instead. In addition, e-handouts were provided to allow students to choose their preferred medium to follow, either the video or the e-handout, which catered to learner differences. Consideration of individual differences and needs has always been a paramount issue for STEM teachers. In the social service course, the method of answering the guiding questions on the worksheet while watching videos also consolidated the content they learned (Azlan et al., 2020).

4.2. Pre- and post-test of emotion and motivation

4.2.1. Analysis of learner empathy

To ensure the consistency of the comparison results, we analyzed the questionnaires of students who completed both the pre- and post-questionnaires. A total of 97 students completed both the pre-questionnaire and the post-questionnaire. The Cronbach's alpha result of the empathy dimension was .85. The paired-sample *t*-test result indicated that there was a significant difference in the learner empathy for pre-test ($M =4.45$, $SD = 0.89$) and post-test ($M =4.80$, $SD = 0.89$); $t(96) = -3.69$, $p < .001$, which demonstrated that students had a significant increase in empathy after completing the project.

4.2.2. Analysis of self-efficacy

The Cronbach's alpha result of the self-efficacy dimension was .91. The paired-sample *t*-test result indicated that there was a significant difference in self-efficacy for pre-test ($M = 3.97$, $SD = 1.07$) and post-test ($M = 4.59$, $SD = 0.98$); $t(96) = -5.45$, $p < .001$, which demonstrated that students had significant increase in self-efficacy after completing the curriculum.

4.2.3. Analysis of interest

The Cronbach's alpha result of the interest dimension was 0.83. The paired-sample *t*-test result indicated that there was a significant difference in self-efficacy for pre-test ($M = 4.20$, $SD = 0.98$) and post-test ($M = 4.62$, $SD = 1.09$); $t(96) = -4.02$, $p < .001$, which demonstrated that students had a significant increase in self-efficacy after completing the project.

The analysis results revealed substantial improvements in students' empathy, self-efficacy, and interest. In the social service course, the first step in product design is to understand the difficulties of potential users. Hence, the improvement of empathy is in line with the original purpose of the course, which is to develop students' empathetic attitudes. The result is consistent with the evidence in the existing literature making connections to social issues strengthens students' empathy (Carlson & Dobson, 2020). In terms of self-efficacy, as mentioned earlier, adequate and timely support played an essential role in enhancing students' self-confidence. Moreover, the smooth advancement of their design under the EDIPT model's guidance also contributed to the enhanced confidence. The result is consistent with Liedtka's (2018) observation, that is, the clear structure of the EDIPT model provided people confidence in innovative design. In terms of interest, the searching for answers to the design problems stimulated self-generated questions and promoted more profound interest (Harackiewicz et al., 2016). There are several advantages in implementing a cross-curricular STEM curriculum design. On the one hand, when students can apply their STEM knowledge to solve real-world social problems, their interest in STEM technical knowledge would be enhanced (Quigley et al., 2019) compared to a STEM-only course. On the other hand, if we have students solve social problems without teaching them sufficiently complex technical knowledge, their solutions could be superficial and less specified.

4.3. Student perception of the curriculum

Interview with the student participants indicated that the transdisciplinary curriculum influenced their empathy, transdisciplinary knowledge, creativity, and willingness to learn. They also expressed the need for more in-class interaction. An interesting phenomenon is that even though the group collaboration mainly happened after class, students expressed that they enjoyed collaborating with teammates. Pseudo names are used in the report to protect the participants' identities.

In response to the question "What is your overall feeling about the curriculum?" students commented:

> Carrie stated that, "*It helped us understand the needs of the people in the community. For example, previously, we knew that people with visual impairment needed help, but we did not know their specific needs. We now have a better understanding of their needs after taking this course*." [Empathy; transdisciplinary knowledge]

> Jerry commented that, "*It stimulated us to observe more details in our daily life. I pay more attention to people in the street to see if anyone needs help. If anyone needs help, I would go ahead and help them out. For example, if something falls out of one's grocery bag, I would pick it up for him or her*." [Empathy]

In response to the question "Which part of the curriculum do you like best? Why?" students commented:

> Henry responded that, "*I like the teamwork part the most, especially the ideation of the product. Because we generated the product idea on our own, and we had lots of discussions on the feasibility and usefulness of the product. We also communicated a lot on how to implement it. The process brings us lots of fun.*" [Collaboration; ideation; fun]

> Jack expressed that, "*Through this learning, we have gained a deeper understanding of STEM. In the meantime, the teamwork brought us together. We came up with our design ideas together, so it made us feel the activity was interesting and meaningful*." [Transdisciplinary knowledge; ideation, collaboration]

Kevin expressed that, "*By doing this project, my creativity improved. In order to complete this project, we came up with many different ideas*." [Creativity]

Regarding the questions "Which part of the curriculum do you think needs improvement? Why?" and "Do you expect to receive any extra support?", most students expressed that they were satisfied with the support and abundant learning resources available at the school, including learning videos and e-handouts. One student made the following suggestions:

Martin suggested, "*In addition to writing proposal, verbal expression is also important. I hope I can have more opportunities to share my ideas verbally, because after speaking out, I will get feedback from peers. Even if it is a critique, you will know which idea is reasonable and which is wrong. Thus, it is a good learning opportunity for us. In addition, expressing our opinions is a good chance to practice our oral presentation skills*." [In-class interaction; verbal expression]

Students also made another suggestion. They expressed that they hoped there could be chances to learn more about STEM and social services. They understood it might be hard to arrange it in class, but perceived it helpful if some self-directed learning resources could be provided in future courses. They expressed that it would be acceptable if teacher guidance is provided occasionally rather than all the time.

In general, students' interview results were consistent with the survey results. The students developed a sense of empathy and were more willing to help others. Students also perceived gaining a more concrete understanding of users' needs, which could be an indicator of enhanced self-efficacy. The students reported continuous intentions to help others and learn more after the courses ended. This result showed that they developed an interest in learning the topic. These findings are quite satisfactory in comparison to several other studies that reported lower student interest in learning in the COVID-19 setting due to a lack of classroom interaction (Ewing & Cooper, 2021).

5. Conclusion and recommendations

Despite the myriad challenges and obstacles facing educators as a direct consequence of the COVID-19 pandemic, there are still opportunities to improve student knowledge by deploying efficient and practical approaches. This study adopted the EDIPT model as the theoretical framework for designing the transdisciplinary social-scientific curriculum. Responding to the constraints caused by the epidemic, multi-model video-facilitated learning approaches were used to organize the classroom activities. While video-based flipped learning is well established to support secondary school subject-based learning (e.g., Jong, 2017; Jong et al., 2019; Lo & Hew, 2021), this study demonstrates that it can also facilitate transdisciplinary STEM learning using the proposed approach. Overall, the video-facilitated transdisciplinary STEM curriculum led to positive changes in students' factual knowledge, design competence, empathy, self-efficacy, and interest.

The curriculum design and the strategic adaptations in the COVID-19 context have meaningful implications for a smooth implementation of STEM teaching during post-pandemic recovery. Firstly, it addressed the gap of having a distinct lack of research in the literature on the crossover of social service and STEM disciplines. The integration of social service, writing, and STEM to develop transdisciplinary skills in secondary school students is a novelty of our work. Due to the complexities of conceptualization, administration, and implementation, interdisciplinary integration is not yet a well-learned field (Cheng & So, 2020). Few studies have explored the pedagogical integration of three or more disciplines. Besides, previous studies have rarely emphasized the importance of designing products for people in need in the community, such as the visually impaired. As this study shows, transdisciplinary STEM teaching and learning can be a way out to empower learners with design competence and transdisciplinary knowledge. Secondly, it also provided evidence on the effectiveness of video-based innovation in supporting learning. The video-based innovation (1) ensured that students were able to successfully carry out STEM learning with minimal disruption when regular STEM in-class time was heavily curtailed; (2) facilitated a connection between students and the real-world context, which enabled them to understand the users' needs better; (3) catered to learner diversity by allowing students to pause or replay hands-on sessions according to their own progress, which consequently increased their confidence in completing more technical challenges.

The COVID-19 is now in its fourth wave, and the fluctuating situation compels us to be flexible in dealing with the new norm. Some of the lessons learned in the study can inform STEM course design irrespective of an epidemic or normal context. Drawing on experience accrued from the study, we have the following recommendations for future STEM curriculum design:

- Put introductory information and conceptual knowledge in pre-class videos to save class time for more challenging issues.
- Use in-class videos in hands-on sections to offset difficulty levels for students, and accompany the videos with e-handouts to allow students to choose their preferred medium.
- Build up a knowledge foundation for students before engaging them in designing solutions to real-life problems.
- Provide a gateway to understanding users' needs through field visits or video-based field visits.
- Arrange on-site supporters for students in the hands-on sessions to provide timely feedback.
- Consider a new collaboration model, i.e., one teacher responsible for live broadcast and the others responsible for on-site support, to alleviate the increased workload in adaption to the emergent situation.

6. Limitation and future research

One of the limitations of this study is that the classroom design should have included more student interactions. Due to the demands of social distancing and short class time, we limited class interactions. When students collaborated on group assignments, they discussed them after class through instant communication tools, e.g., WhatsApp® or Zoom®, without any teacher or mentor involvement. During the limited in-class discussion time, the teachers explained the group work requirements and scaffolded their discussion with prepared worksheets, but did not model any answers to give students more control over their projects and encourage creative thinking (van Leeuwen & Janssen, 2019). If time permits, more student-to-student interactive activities such as peer sharing could be introduced inside the classroom to make the course more engaging and allow students to learn more through peer interaction. Additionally, it would be interesting to explore if other strategies can be implemented to enhance in-class engagement. Video-based field visits have proven to offer specific advantages. For example, they can reduce the financial cost and staffing needed to organize large events (Chang et al., 2020; Jong et al., 2020). It has more flexibility in terms of time, as teachers can show it in class at any time. If conditions permit, it would be more beneficial to have students visit social service organizations in person, which is even more impactful than the video-based field visit (Friess, 2016). Due to the unique context, this study conducted a single group pre- and post-test experiment. In the future, a comparison group can be included. In general, it is an exciting exploration to connect social service with STEM education. Further research in this direction would be imperative.

Acknowledgement

The work described in this paper was substantially supported by The Hong Kong Jockey Club Charities Trust (Project Title: Jockey Club Community Care and STEM in Action [Project S/N Ref: JC 2019/0112]).

References

Atman, C. J., Adams, R. S., Cardella, M. E., Turns, J., Mosborg, S., & Saleem, J. (2007). Engineering design processes: A Comparison of students and expert practitioners. *Journal of Engineering Education, 96*(4), 359-379.

Azlan, C. A., Wong, J. H. D., Tan, L. K., Huri, M. S. N. A. D., Ung, N. M., Pallath, V., Tan, C. P. L., Yeong, C. H., & Ng, K. H. (2020). Teaching and learning of postgraduate medical physics using Internet-based e-learning during the COVID-19 pandemic–A Case study from Malaysia. *Physica Medica, 80*, 10-16.

Bates, A. W. (2019). *Teaching in a digital age* (2nd ed.). Tony Bates Associates Ltd. https://pressbooks.bccampus.ca/teachinginadigitalagev2/

Bond, M. (2020). Facilitating student engagement through the flipped learning approach in K-12: A Systematic review. *Computers & Education, 151*, 103819. https://doi.org/10.1016/j.compedu.2020.103819

Bozkurt, A., Jung, I., Xiao, J., Vladimirschi, V., Schuwer, R., Egorov, G., Lambert, S., Al-Freih, M., Pete, J., Olcott, Jr. D., Rodes, V., Aranciaga, I., Bali, M., Alvarez, A. Jr., Roberts, J., Pazurek, A., Raffaghelli, J. E., Panagiotou, N., de Coëtlogon, P., Shahadu, S., Brown, M., Asino, T. I., Tumwesige, J., Reyes, T. R., Ipenza, E. B., Ossiannilsson, E., Bond, M., Belhamel, K., Irvine, V., Sharma, R. C., Adam, T., Janssen, B., Sklyarova, T., Olcott, N., Ambrosino, A., Lazou, C., Mocquet, B., Mano, M., & Paskevicius, M. (2020). A Global outlook to the interruption of education due to COVID-19 pandemic:

Navigating in a time of uncertainty and crisis. *Asian Journal of Distance Education, 15*(1), 1-126. https://asianjde.com/ojs/index.php/AsianJDE/article/view/462

Brame, C. J. (2016). Effective educational videos: Principles and guidelines for maximizing student learning from video content. *CBE—Life Sciences Education, 15*(4). https://doi.org/10.1187/cbe.16-03-0125

Brecht, H. D., & Ogilby, S. M. (2008). Enabling a comprehensive teaching strategy: Video lectures. *Journal of Information Technology Education: Innovations in Practice, 7*(1), 71-86.

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.

Carlson, D. J., & Dobson, T. (2020). Fostering empathy through an inclusive pedagogy for career creatives. *International Journal of Art & Design Education, 39*(2), 430-444.

Chai, C. S., Rahmawati, Y., & Jong, M. S. Y. (2020). Indonesian science, mathematics, and engineering preservice teachers' experiences in STEM-TPACK design-based learning. *Sustainability, 12*(21). https://doi.org/10.3390/su12219050

Chai, C. S., Wang, X., & Xu, C. (2020). An extended theory of planned behavior for the modelling of Chinese secondary school students' intention to learn artificial intelligence. *Mathematics, 8*(11). https://doi.org/10.3390/math8112089

Chang, S. C., Hsu, T. C., Kuo, W. C., & Jong, M. S. Y. (2020). Effects of applying a VR-based two-tier test strategy to promote elementary students' learning performance in a geology class. *British Journal of Educational Technology, 51*(1), 148–165.

Chen, C. M., & Wu, C. H. (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education, 80,* 108-121.

Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods, 4,* 62-83.

Cheng, Y. C., & So, W. W. M. (2020). Managing STEM learning: A Typology and four models of integration. *International Journal of Educational Management, 34*(6), 1063-1078.

Conlon, M. M., & McIntosh, G. L. (2020). Perceptions of realism in digital scenarios: A Mixed methods descriptive study. *Nurse Education in Practice*, *46*, https://doi.org/10.1016/j.nepr.2020.102794

D'addato, T., & Miller, L. R. (2016). An Inquiry into flipped learning in fourth grade math instruction. *The Canadian Journal of Action Research*, *17*(2), 33-55.

English, L. D. (2016). STEM education K-12: Perspectives on integration. *International Journal of STEM Education, 3,* https://doi.org/10.1186/s40594-016-0036-1

Engzell, P., Frey, A., & Verhagen, M. D. (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences of the United State of America*, *118*(17). https://doi.org/10.1073/pnas.2022376118

Epler-Ruths, C. M., McDonald, S., Pallant, A., & Lee, H. S. (2020). Focus on the notice: Evidence of spatial skills' effect on middle school learning from a computer simulation. *Cognitive Research: Principles and Implications, 5*. https://doi.org/10.1186/s41235-020-00263-0

Ewing, L. A., & Cooper, H. (2021). Technology-enabled remote learning during Covid-19: Perspectives of Australian teachers, students and parents. Technology, *Pedagogy and Education, 30*(1), 39-54.

Fiş Erümit, S. (2020). The distance education process in K–12 schools during the pandemic period: Evaluation of implementations in Turkey from the students' perspective. *Technology, Pedagogy and Education, 30*(1), 71-90. https://doi.org/10.1080/1475939X.2020.1856178

Friess, D. A., Oliver, G. J. H., Quak, M. S., & Lau, A. Y. A. (2016). Incorporating "virtual" and "real world" field trips into introductory geography modules. *Journal of Geography in Higher Education, 40*(4), 546-564.

Geng, J., Jong, M. S. Y., Chai, C. S. (2019). Hong Kong teachers' self-efficacy and concerns about STEM education. *The Asia-Pacific Education Researcher, 28*(1), 35–45.

Gibbs, P. (2015). Transdisciplinarity as epistemology, ontology or principles of practical judgment. In P. Gibbs (Ed.), *Transdisciplinary Professional Learning and Practice* (pp. 151–164). Springer International Publishing.

Gunckel, K. L., & Tolbert, S. (2018). The Imperative to move toward a dimension of care in engineering education. *Journal of Research in Science Teaching, 55*(7), 938-961.

Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An Empirical study of MOOC videos. *Proceedings of the First ACM Conference on Learning@ Scale Conference* (pp. 41–50). Association for Computing Machinery.

Harackiewicz, J. M., Smith, J. L., & Priniski, S. J. (2016). Interest matters: The Importance of promoting interest in education. *Policy Insights from the Behavioral and Brain Sciences, 3*(2), 220-227.

Hasso Plattner Institute of Design. (2010). *An Introduction to design thinking process guide*. https://www.evansville.edu/changemaker/downloads/Process%20Guide%20for%20Design%20Thinking.pdf

Honey, M., Pearson, G., & Schweingruber, H. A. (Eds.). (2014). *STEM integration in K-12 education: Status, prospects, and an agenda for research.* National Academies Press.

Hong, H. Y., Lin, P. Y., Chen, B., & Chen, N. (2019). Integrated STEM learning in an idea-centered knowledge-building environment. *The Asia-Pacific Education Researcher, 28*(1), 63-76.

Hutchison, M. (2016). The Empathy project: Using a project-based learning assignment to increase first-year college students' comfort with interdisciplinarity. *Interdisciplinary Journal of Problem-Based Learning*, 10(1). https://doi.org/10.7771/1541-5015.1580

Jong, M. S. Y. (2017). Empowering students in the process of social inquiry learning through flipping the classroom. *Educational Technology & Society, 20*(1), 306–322.

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

Jong, M. S. Y. (2019b). To flip or not to flip: Social science faculty members' concerns about flipping the classroom. *Journal of Computing in Higher Education, 31*(2), 391–407.

Jong, M. S. Y., Chen, G. W., Tam, V., & Chai, C. S. (2019). Adoption of flipped learning in social humanities education: The FIBER experience in secondary schools. *Interactive Learning Environments, 27*(8), 1222–1238.

Jong, M. S. Y., Tsai, C. C., Xie, H., & Wong, F. K. K (2020). Integrating interactive learner-immersed video-based virtual reality into learning and teaching of physical geography. *British Journal of Educational Technology, 51*(6), 2064-2079.

Lapitan, L., Jr., Tiangco, C. E., Sumalinog, D. A. G., Sabarillo, N. S., & Diaz, J. M. (2021). An Effective blended online teaching and learning strategy during the COVID-19 pandemic. *Education for Chemical Engineers, 35*, 116-131.

Lee, O., & Campbell, T. (2020). What science and STEM teachers can learn from COVID-19: Harnessing data science and computer science through the convergence of multiple STEM subjects. *Journal of Science Teacher Education, 31*(8), 932-944.

Liedtka, J. (2018). Why design thinking works. *Harvard Business Review*, *96*(5), 72-79.

Lin, Y. T., & Chen, C. M. (2019). Improving effectiveness of learners' review of video lectures by using an attention-based video lecture review mechanism based on brainwave signals. *Interactive Learning Environments, 27*(1), 86-102.

Linnenbrink-Garcia, L., Patall, E. A., & Pekrun, R. (2016). Adaptive motivation and emotion in education: Research and principles for instructional design. *Policy Insights from the Behavioral and Brain Sciences, 3*(2), 228–236.

Lo, C. K., & Hew, K. F. (2021). Developing a flipped learning approach to support student engagement: A Design-based research of secondary school mathematics teaching. *Journal of Computer Assisted Learning, 37*(1), 142-157.

Luo, Z., Dang, Y., & Xu, W. (2019). Academic interest scale for adolescents: development, validation, and measurement invariance with Chinese students. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.02301

Maiorca, C., Roberts, T., Jackson, C., Bush, S., Delaney, A., Mohr-Schroeder, M. J., & Soledad, S. Y. (2021). Informal learning environments and impact on interest in STEM careers. *International Journal of Science and Mathematics Education,19,* 45-64.

Morrin, A. M., & Liston, M. (2020). Engaging children with authentic STEAM learning experiences through design-based approaches. *Connected Science Learning, 2*(4). https://www.nsta.org/connected-science-learning/connected-science-learning-october-december-2020/engaging-children

Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research, 62*(3), 229-258.

Organization for Economic Cooperation and Development. (2019). *PISA 2021 creative thinking framework (third draft)* https://www.oecd.org/pisa/publications/PISA-2021-creative-thinking-framework.pdf

Pal, D., & Patra, S. (2020). University students' perception of video-based learning in times of COVID-19: A TAM/TTF perspective. *International Journal of Human–Computer Interaction.* https://doi.org/10.1080/10447318.2020.1848164

Quigley, C. F., Herro, D., & Baker, A. (2019). Moving toward transdisciplinary instruction: A Longitudinal examination of STEAM teaching practices. In M. Khine & S. Areepattamannil (Eds.), *STEAM Education* (pp. 143-164). Springer

Rulifson, G., & Bielefeldt, A. (2017). Motivations to leave engineering: Through a lens of social responsibility. *Engineering Studies, 9*(3), 222-248.

Simeon, M. I., Samsudin, M. A. & Yakob, N. (2020). Effect of design thinking approach on students' achievement in some selected physics concepts in the context of STEM learning. *International Journal of Technology and Design Education* https://doi.org/10.1007/s10798-020-09601-1

So, H. J., Jong, M. S Y. & Liu, C. C. (2020). Computational thinking education in the Asian pacific region. *The Asia-Pacific Education Researcher, 29*. https://doi.org/10.1007/s40299-019-00494-w

Song, Y., Jong, M. S. Y., Chang, M., & Chen, W. (2017). "HOW" to design, implement and evaluate the flipped classroom? – A Synthesis. *Educational Technology & Society, 20*(1), 180–183.

STEM Task Force. (2014). *Innovate: A blueprint for science, technology, engineering, and mathematics in California public education.* Californians Dedicated to Education Foundation.

Tabakin, A. L., Patel, H. V., & Singer, E. A. (2021). Lessons learned from the COVID-19 pandemic: A Call for a national video-based curriculum for urology residents. *Journal of Surgical Education, 78*(1), 324-326.

Tsai, M. N., Liao, Y. F., Chang, Y. L., & Chen, H. C. (2020). A Brainstorming flipped classroom approach for improving students' learning performance, motivation, teacher-student interaction and creativity in a civics education class. *Thinking Skills and Creativity, 38*. https://doi.org/10.1016/j.tsc.2020.100747

van Leeuwen, A., & Janssen, J. (2019). A Systematic review of teacher guidance during collaborative learning in primary and secondary education. *Educational Research Review, 27*, 71-89.

Vasquez, J., Sneider, C., & Comer, M. (2013). *STEM lesson essentials, grades 3–8: Integrating science, technology, engineering, and mathematics.* Heinemann.

Vossen, H. G. M., Piotrowski, J., Valkenburg, P. M. (2015). Development of the adolescent measure of empathy and sympathy (AMES). *Personality and Individual Differences, 4*, 66-71.

Yates, A., Starkey, L., Egerton, B., & Flueggen, F. (2020). High school students' experience of online learning during Covid-19: The influence of technology and pedagogy. *Technology, Pedagogy and Education, 30*(1), 55-69. https://doi.org/10.1080/1475939X.2020.1854337

Appendix I. The video-facilitated strategies in the COVID-19 context

	Challenges	Video-facilitated strategies	Other administrative support
STEM	• Shortened class time; not all knowledge can be taught in class. • No group work in the classroom, which makes learning more challenging for individual students. • Individual differences in students' understanding and operating proficiency.	• Pre-class: Provide less than 3-minute short videos to introduce the introductory concepts. • In-class: Pre-recorded all hands-on sections. Students can follow the video in the hand-on sections. The videos are accompanied by e-handouts.	• To alleviate the workload faced by teachers in adapting to the adjusted instructional content and approach, the STEM teaching team made adjustments to their collaboration approach. One teacher gave lectures through live broadcast, while one teacher and one student mentor provided on-site support in every classroom. Students could seek help directly from the on-site mentor if they had technical problems during the hands-on session.
Social service	• There is no field visit, so students' understanding of the users' needs may be vague. • The direct instruction format may not engage students. • The class time is shortened, and it is impossible to cover	• In-class: Video-based field visit. The project team visited the social service organizations prior to the course, and interviewed the potential users and the staff working in the social service organizations. Students could watch a video to learn about the potential users' difficulties in daily lives and identify	• Similarly, for the social service course, one teacher gave lectures to four classes simultaneously via live broadcast, and one teacher in each classroom guided students to complete worksheets and discussions.

	all the contents in the class.	their needs. • While watching the video, students would answer the guiding questions on the worksheet to record the users' essential needs. • Post-class: Students could watch the extracurricular extension videos on their own. Students could watch the extracurricular extension videos on their own. The videos introduced some new high-tech products that can help people in need.	
Proposal writing	• In the format of direct instruction, students may have a vague understanding of the product development background. Students may not have a clear understanding of how to introduce a product in the form of a proposal.	• In-class: Showed selected exemplars of technological products and instructed students to complete the proposal worksheet for a sample product.	• Lessons were taught by the language teacher of each class.

Rof, A., Bikfalvi, A., & Marques, P. (2022). Pandemic-accelerated Digital Transformation of a Born Digital Higher Education Institution: Towards a Customized Multimode Learning Strategy. *Educational Technology & Society*, *25* (1), 124-141.

Pandemic-accelerated Digital Transformation of a Born Digital Higher Education Institution: Towards a Customized Multimode Learning Strategy

Albert Rof*, Andrea Bikfalvi and Pilar Marques
University of Girona, Spain // u1008518@campus.udg.edu // andrea.bikfalvi@udg.edu // pilar.marques@udg.edu
*Corresponding author

ABSTRACT: The COVID-19 pandemic has forced the digitalization of the majority of universities, prior to which they were largely operating using face-to-face modes of learning. Increased competition in the digital environment places universities under greater pressure to offer an innovative learning experience. The purpose of this paper is to understand the effects of the sudden pandemic on the ongoing process of digital transformation (DT) and how the learning value proposition of higher education institutions (HEIs) has been affected. The research is based on a single case study of a born digital university, focusing on the changes made to the learning value proposition, and particularly to the multimode learning offer. The paper uncovers the relation between multimodality and customized and personalized learning, all of which are dependent on the use of digital educational technology. The originality of this paper is its longitudinal look at a single case, observing how the significant DT process already underway prior to the pandemic has been impacted by it, accelerating the process, and clarifying the envisaged post-pandemic future for HEIs. Another distinctive aspect is the consideration of the learning proposition as a core element and part of a larger and interdependent value proposition within the overall HEIs business model.

Keywords: Higher Education Institutions, Customized learning, Multimode learning, Digital transformation, Business model

1. Introduction

The impact of the COVID-19 pandemic on higher education institutions (HEIs), defined as universities, colleges, and polytechnics that offer degrees beyond secondary education, has been dramatic on a global scale. The so-called emergency or forced digitalization allowed HEIs to continue offering their students learning opportunities when social distancing and lockdown were mandatory. The COVID-19 shock has been revolutionary and has impacted the entire higher education system, causing a drastic shift in the scale of change (Alvesson & Sveningsson, 2015) in a sector that was already immersed in a continuous digitalization process, with digital technologies threatening to disrupt HEIs (Posselt et al., 2018).

Research carried out prior to the pandemic already considered the adoption of digital technologies and processes inevitable to remain a relevant player in higher education (Khalid et al., 2018). Most HEIs were already proving to be adaptive to these technologies, implementing new teaching and learning methodologies rapidly, at least operationally (Alvesson & Sveningsson, 2015). According to a global survey of 424 HEIs in 106 countries (Marinoni et al., 2020), at the onset of the pandemic, 67% of HEIs were able to replace classroom teaching with online distance teaching and learning. This research concludes that the forced learning and testing of new digital tools and methodologies (e.g., video conferences) has changed the digital mentality of teachers, opening a window to explore more flexible learning paths now that online learning is envisaged to be a more integral part of teaching plans. What remains to be seen is whether this proven operational capacity to change and adapt to an emergency situation will become fully integrated into HEIs and evolve into a strategic capacity to implement change (Alvesson & Sveningsson, 2015). This integration will be essential in a sector whose boundaries are being aggressively trespassed by new competitors, including the so-called “EdTech” companies, understood as companies that intensively apply “technological resources and processes for learning and teaching purposes” (Kaplan, 2020). These new entrants are competing with innovated-digitalized business models to change the rules of the training industry (Posselt et al., 2018).

Marinoni et al. (2020) uncover that the pandemic has significantly helped increase inequality in learning opportunities, at least in the short run, since almost a third of HEIs did not adapt fast enough to the new digitalization-forced reality. Although this situation is expected to be resolved in the near future, it reminds us the challenges arising from the previously acknowledged academic digital gap (Bond et al., 2018).

The COVID-19 pandemic is an example of an exogeneous shock, defined as “a period of prolonged and widespread crisis in which actors struggle to reconstitute all aspects of social life” (Fligstein & McAdam, 2011, p. 32). The impact of exogenous shocks has been explored at the business model level (Corbo et al., 2018; Morgan et al., 2020), and the specific impact of COVID-19 has been explored in other contexts such as start-ups (Kuckertz et al., 2020) and family businesses (Soluk et al., 2021), among others. Research on the impact of COVID-19 on the HEI sector has also been carried out (Marinoni et al., 2020; Vlachopoulos, 2020), but what is still unknown is its impact on the business model of HEIs, especially in terms of the effect on the DT process already underway. The success and direction of the DT of HEIs in the midst and aftermath of COVID-19 is of present importance because HEIs are a backbone for training, knowledge generation and transfer, and ultimately social development. Beyond the COVID-19 impact, the findings of this research can also be informative for future shocks to the HEI sector.

HEIs are being forced to adapt to the ongoing cultural and societal changes challenging traditional educational practices, a central aspect of which is the rapid and continual development of digital technologies, some of which have been specifically developed for educational purposes. Education research should be grounded within current social, political, and philosophical changes, with a strong call towards sustainability (Stepanyan et al., 2013). Building on the societal issue of technological change related to education, we aim to contribute to the debate on the present and future of a higher education immersed in a continuous DT, exposed to a highly competitive landscape, and affected by exogeneous shocks of a societal, health, economic, and sectoral nature. Most scholarly approaches to higher education, educational technology, and the business models of HEIs tend to focus on dual associations, mostly higher education and educational technology, with little research at the intersection of the three issues. The research at this intersection also responds to calls for further enquiry into new business models based on technological innovations (Stepanyan et al., 2013), especially when they encompass mobile, ubiquitous, and game-based learning (Kinshuk et al., 2013), the cultural diversity of stakeholders when deploying technology-assisted learning in international contexts (Habib et al., 2014), and the issue of inequality concerning the web lecturing mode (Montrieux et al., 2015), among others. Additionally, in the context of the ongoing digitalization process, there have been recent calls for further research into different aspects of customized or personalized learning (Lee et al., 2018) in higher education, including the challenges of digitalisation in different learning contexts and student engagement and motivation within these personalized learning environments (Alamri et al., 2021). Other authors call for more research on personalized learning content and delivery modes (Xie et al., 2019), the performance of technology platforms, and personal learner profiles (Alamri et al., 2021), among others.

To address the knowledge gap of the impact of COVID-19 on the future of the HEI sector, the purpose of this paper is to understand the effect of the unexpected pandemic on the learning value proposition of HEIs as a core element of their business models (BM), adding to the already huge impact of the ongoing DT. We thus propose the following research question: How has the COVID-19 shock affected the ongoing DT of HEIs, especially as regards the learning value proposition?

We use a longitudinal single case study to investigate the research question, observing how the significant DT process is currently being impacted by the pandemic, accelerating the desired vision of the studied HEI. An original element of this study is its positioning at the triple intersection of COVID-19, digital transformation, and business models in the HEI sector. Our theoretical framework and empirical findings uncover the use of multimodality to facilitate customized and personalized learning. We build on existing research to explore multimodality in teaching, mainly from two approaches. The first is the taxonomic proposal of Margulieux et al. (2016), which is based on three dimensions, face-to-face versus online learning, the delivery medium, and the instruction type. And the second focuses on the main e-learning forms in higher education, namely distance, formal, and open education (Nguyen et al. 2019), and online distance learning (Kaplan & Haenlein, 2016), understood as all forms of instruction where the student is separated by distance from the instruction and whose interactions are mediated by digital technologies. Within this context, this paper understands the concept of multimode digital learning as the matrix of digital methods, forms, and tools, including direct instruction via synchronous video conferences and asynchronous videos, group-project-based learning, and online exams, that can be used for digital or digitally enhanced learning. In this paper we argue how this matrix will allow HEIs and students to respectively offer and choose from a very large set of learning combinations, which will eventually lead to HEIs offering a customized learning value proposition that will change the *what, when, how,* and *where* of the learning journey.

This introduction is followed by a theoretical section that sets the frame for the research and identifies the gap. The methods section describes the case study chosen and the methodological process followed. The results section presents the empirical findings, evidenced by interviewees’ direct quotations and structured codifications of the changes in the learning value proposition. Next, the discussion considers the results in the light of the research question and the theoretical background. Last, a concluding section provides an overall assessment of the paper with its highlighted contribution, some limitations, and future research proposals.

2. Theoretical background

2.1. Educational technology, learning multimodality, and personalized learning

2.1.1. Educational technology

It can be argued that research on educational technology has not generally been supported by and connected to learning theories. There are, however, some attempts to do so and connections have been made with existing theories, including constructivism and behaviourism (Albirini, 2007). Behaviourism considers learning as a reactive process (Clark & Salomon, 1986), with students taking a passive role and a teacher-centric lecturing approach (Gärdenfors & Johansson, 2005). Educational technology within the digital milieu, however, does not fit well with behaviourism since digital means enabling an active and more student-centric approach more in line with a constructivist view (Albirini, 2007). The constructivist theory pioneered by Jerome Brunner in 1966 (Sejzi & Aris, 2012) proposes that "learning is an active process where students construct knowledge or new concepts based on their experiences" (Alamri et al., 2021, p. 427), becoming autonomous and independent learners (Alamri et al., 2021) who take responsibility for their learning anytime and anywhere (Sejzi & Aris, 2012). Information and communication technologies such as learning management systems and videoconferencing tools, among others, can provide a constructivist context for learning (Sejzi & Aris, 2012), even if there are some concerns about the lack of clarity as to what students are constructing (Gärdenfors & Johansson, 2005).

2.1.2. Learning multimodality

Extant research underlines the importance of DT in HEIs (Kaplan & Haenlein, 2016) and its impacts on different processes and groups, including students, staff, and professors. A myriad of digitally-driven opportunities are explored, including adding digital technologies to be able to develop new learning strategies that are more interactive and based on co-learning (Pucciarelli & Kaplan, 2016), and customising individual lessons (Renz & Hilbig, 2020). With the rise of new teaching and learning methods that integrate new digital technologies, including artificial intelligence, machine learning, and learning analytics, the HEIs' BM is becoming more digitalized and data-based (Renz & Hilbig, 2020). This digitalization of HEIs opens a world of options, including digital/non-digital hybridizations of learning systems and tools that increase its multimodality.

There are some attempts to define the multimode teaching options that emerge from combining face-to-face and online learning (Margulieux et al., 2016), including hybrid, blended, flipped, and inverted methodologies, among others. The taxonomy has been established by combining two dimensions, the delivery medium (via an instructor and/or via technology, when an electronic system mediates between the teacher and the learner) and the instruction type (if students are mainly receiving content during instruction and/or applying content). This combination of teaching modes contributes to adapting to the personal preferences and type of learner (Felder & Silverman, 1988), e.g., visual or verbal, active or reflective, and so on.

As regards technological means, some previous research has focused on the main e-learning forms in higher education, including distance, formal (homologated), and open education (Nguyen et al., 2019). According to the same authors, e-learning represents a new way of teaching and learning which is: (i) more learner-centric and learner-personalized, (ii) supported by the ever changing digital technologies that offer ubiquity in the access and delivery of teaching resources and services anytime, anywhere; and (iii) uses interactive, collaborative, and personalized modes.

Other authors understand online distance learning (Kaplan & Haenlein, 2016) as all forms of instruction when the student is separated by distance from the instructor and when interactions are mediated by digital technologies. Distance learning can be developed with time separation (asynchronous) or not (synchronous). Considering that the number or participants can be limited or unlimited, this time and space combination offers an interesting multimode portfolio of teaching opportunities for distance learning. For example, the asynchronous method allows for Massive Open Online Courses (MOOCs; open-access online courses for the open community) and Small Private Online Courses (SPOCs; for limited participants), while the synchronous method allows for Synchronous Massive Online Courses (SMOC; open access but with students simultaneously digitally present) and Synchronous Small Online Courses (SSOC; the same as SMOCs but for a limited number of participants). While all these possibilities already existed pre-pandemic, their application was uneven and optional. HEIs embraced the former innovations at their own pace and under the influence of various contextual, organizational, and individual factors.

2.1.3. Personalized learning

Multimodality opens a myriad of possibilities to offer learning experiences more adapted to students' needs and wishes. Despite increased interest in personalized learning at the academic level in recent years, there is no agreed definition of the concept (Shemshack & Spector, 2020; Schmid & Petko, 2019). A recent systematic review of published research on personalized learning has revealed that different terms, such as adaptive learning, individualized instruction, and customized learning, have been used interchangeably (Shemshack & Spector, 2020). Customized learning considers "individual differences and needs, characteristics, interests, and academic mastery" (Shemshack & Spector, 2020, p. 6). According to Hsieh and Chen (2016) personalized learning aims to match the learning experience with the needs of different cognitive style groups, using adaptivity to automatically tailor content, structure, and presentation to each individual (Treiblmaier et al., 2004). Personalized learning is controlled by the system, or the educational technology platforms, and it is system driven (Kay, 2001). In contrast, customized learning aims to tailor the experience to the needs of each individual, endowing individuals with adaptability to make modifications to the content presentation and format layout by themselves (Treiblmaier et al., 2004). Customized learning is controlled by the user (Hsieh & Chen, 2016), so it is user driven (Kay, 2001), with users involved in the initiation, proposal, selection, and even production of learning elements (Kobsa et al., 2001). Users can choose from a menu of available options (Frias-Martinez et al., 2009) that offer different degrees of customization (Teng, 2010), reducing the risk of improper adaptation (Findlater & McGrenere, 2004) of personalized systems. Customization and personalization can both be applied to accommodate the diversity of students' cognitive styles (Hsieh & Chen, 2016).

From the perspective of learning theory, personalized learning is ingrained in the constructivist theory (Alamri et al., 2021), and has the potential to develop learner-centred strategies, with information technology platforms facilitating this process (Albirini, 2007). However, customized learning involves more agency from the student, which is even more aligned with a constructivist view.

2.2. Digital transformation and the HEI business model

While there is no unified definition of digital transformation (DT), a recent review of 124 articles has defined the concept as "a fundamental change process enabled by the innovative use of digital technologies, accompanied by the strategic leverage of key resources and capabilities aimed at radically improving an entity (an organization, a business network, an industry, or society) and redefining its value proposition for its stakeholders" (Gong & Ribiere, 2021, p. 12).

The concept of business model has been widely studied in academia (Foss & Saebi, 2017) and much used in the business world, especially in entrepreneurial environments thanks to the popularization of tools like the Business Model Canvas (Osterwalder et al., 2010). A generally accepted definition of BM describes the concept as a "story" that essentially explains how firms work (Magretta, 2002) and how a firm does business (Demil et al., 2015), or "the rationale of how an organization creates, delivers and captures value" (Osterwalder et al., 2010). When the focus is on explaining the different elements or dimensions that configure the BM, there are different BM frameworks such as the BM in five value dimensions, namely value proposition, value communication, value creation, value delivery, and value capture (Abdelkafi et al., 2013).

Rising consensus that business practices are becoming necessary in HEIs (Pucciarelli & Kaplan, 2016) has led to the recent use of the BM concept and approach in the context of universities (Abdelkafi et al., 2018; Rosi et al., 2018). Posselt et al. (2018) analyse the evolution of universities towards being more entrepreneurial, pointing to the importance for universities of expanding and digitalizing their offering. Only limited research has explored how the business model is innovated due to the impact of DT in the particular context of HEIs (Rof et al., 2020).

Digitalization is changing the higher education sector. New "EdTech" companies are entering the sector with innovative business models (Kaplan, 2020), some of them integrating state of the art technologies for education purposes, including learning analytics and artificial intelligence, into their BM (Renz & Hilbig, 2020). Furthermore, recent research states that digital technologies are disrupting universities (Posselt et al., 2018) and that HEIs must adapt to technological changes if they want to stay relevant (Zulfikar et al., 2018; Khalid et al., 2018). The growth of distance learning and derivative formats (MOOCs, social media, etc.) can potentially remodel the education industry in the near future, increasing the risk of disappearance of the non-adapted players (Kaplan & Haenlein, 2016). In the same line, some argue that implementing new technologies is essential to be digitally relevant, and that the real challenge is the appropriate execution of digital plans and strategies (Nguyen, 2018). More particularly, other research explores how DT impacts professors and students, including how to address the academic digital gap by developing professors' digital skills since students are very motivated to use digital tools for learning (Bond et al, 2018).

2.3. The effects of COVID-19 on the HEI value proposition

The COVID-19 shock has been explored in other contexts such as start-ups (Kuckertz et al., 2020) and family businesses (Soluk et al., 2021). Recent research has also focused on the impact of COVID-19 on HEIs, covering multiple topics as diverse as whether online education should be considered a threat or an opportunity (Vlachopoulos, 2020), how digital innovation was encouraged during the emergency (Agasisti et al., 2020), and how cloud services can support online learning (Bhardwaj et al., 2021). As regards teaching modalities, there are studies on how online teaching methodologies such as the inverted classroom (flipped) can add value in the new context (Izagirre-Olaizola & Morandeira-Arca, 2020), how examination issues have been resolved creatively by replacing exams with research papers (El-Bassiouny & Mohamed, 2020), and what learning strategies were attempted in the initial stage of pandemic and what results they produced (Dietrich et al., 2020), among others. Current research is focusing on the situation post the initial stage of the COVID-19 pandemic, the so-called "new normality" (Nandy et al., 2021; Tesar, 2020).

Despite the research gap on the impact of the pandemic on the HEI business model, research on the impact of COVID-19 on HEIs (Marinoni et al., 2020) is showing that the forced shift from face-to-face teaching to online distance teaching and learning methods has created both challenges and opportunities that impact to varying degrees on the different blocks of the BM. For example, a forced digitalisation has been triggered (Marinoni et al., 2020), causing a change in the learning value proposition, or the bundle of teaching products and services offered by the HEI, and creating an opportunity to make the future higher education sector more flexible. The increased use of multimode learning approaches, such as blending face-to-face and online learning activities (hybrid learning), and combining synchronous learning with asynchronous learning, are among these opportunities. All these new modes change the nature of the relationships and channels used with students, modifying the value the student receives from the HEI through a transformed learning value proposition. As regards the teaching staff, the forced learning and testing of new digital tools and methodologies (e.g., video conferences) has changed their digital mentality, which is expected to influence future teaching plans (part of the value proposition) to make online learning more integral, triggering innovation in both pedagogical methodologies (e.g., examinations) and delivery modalities. Other relevant identified opportunities include investing in cloud services to digitalize access to resources (e.g., library) and processes (e.g., administrative procedures), more remote working opportunities for lecturers and staff, and increased awareness among students of lifelong learning opportunities.

All these changes, which have already been applied to adapt teaching to the state of emergency, impact the current learning value proposition, a core element of BM. There will also be further repercussions of different types on most of the blocks that configure the business model. Understanding this configuration calls for a detailed analysis of multimode teaching/learning and how it affects the elements of BM building blocks.

3. Method

To answer the research question as to how the COVID-19 shock affects the DT of HEIs, this paper looks longitudinally at a single case study of a pioneering, born digital HEI headquartered in Spain. Qualitative in approach, the research design observes the studied HEI in two separate moments in time, a year before the start of the COVID-19 pandemic (November 2018 to January 2019) and a short time after its emergence (July 2020-December 2020), to understand how the significant DT process started before the pandemic is being impacted by it, and how the process is accelerating the desired vision of the studied HEI. Case studies provide qualitative and rich data and allow the study of current management challenges (Yin, 2009). The shock effect of COVID-19 triggering the so-called “forced digitalization” adds complexity to a DT process that was already impacting the business model of the HEI. The complexity and depth of the combined impacts of COVID-19/DT make the use of a single case suitable to observe in depth the experiences and insights of its participants regarding DT and its impact on the BM both before and after the emergence of the COVID-19 pandemic, and particularly on the online learning value proposition.

Table 1. Methodological summary and interviewee

Methodological orientation	Qualitative exploratory research discourse analysis				
Technique	Case study				
Number of cases	One				
Field work	Ex-ante (before the emergence of the pandemic); interviews from Nov 2018 to Jan 2019. Secondary data: Oct 2018 to Jan 2019 Ex-post (after the emergence of the pandemic): administered questionnaire from Jul 2020-Dec 2020. Secondary data: Jul 2020 to Dec 2020				
Primary source of information	Individual interviews				
Participant selection	Purposive sampling Executive committee members, executive positions Criteria: heterogeneity by function, position, contractual relationship E-mail approach				
Instrument used	Semi-structured questionnaires				
Main topics of the interview	Pre-pandemic: Digital transformation concept (DT). Impact of DT. Main DT innovations. Main challenges and opportunities derived from DT. Tensions derived from the DT process, and solutions Post-pandemic: areas of the university most significantly impacted by the effects of COVID-19 forced digitization, worst and best situations and how they were handled, impact on the vision of what DT is and its importance, impact map for stakeholders, challenges and opportunities, and visions of the future because of the impacts of DT and COVID-19				
Setting and data collection	Pre-pandemic: Interviews conducted in the workplace. Interview guide provided in advance. Audio recording. Field notes by authors during and after interviews. Additional/missing/incomplete information requested after the interviews Post-pandemic: administered questionnaire post-pandemic				
Data analysis	2 coders Coding: Primary codes—Themes; Secondary codes—Sub-topics; Aggregate dimensions Themes derived from the data				
Secondary sources of information	Public data: website, annual reports, HEI presentations, press news				
Number of informants	4	1	1	1	1
Informants work position	Total	DMO	VRSPR	VRCE	VPOT
Function		Innovation projects (Admin., teaching, research)	Strategy and Research	Competitiveness and Employability	Operations and Technology
Background		Comp. Engineering	Medicine and Surgery	Economics, Finance.	ICT
Duration of interview (minutes)	323′	73′	118′	55′	77′

Notes. HEI, higher education institution. DMO (Management): Director Management Office; VRSPR (Strategy): Vice Rector of Strategic Planning and Research; VRCE (Competitiveness): Vice Rector of Competitiveness and Employability; VPOT (Operations and Technology): Vice President of Operations and Technology.

Table 1 presents a methodological summary and provides details of the participants interviewed, including their current function at the institution and their background. The selection criteria included people who altogether represented a variety of functions (innovation policy, strategy and research, competitiveness and employability, operations, and technology) and positions occupied (vice-rectors, vice-presidents), and who had a consolidated tenure in the HEI (average of 12 years in the HEI and 2.5 years in the current position). The single case selected is a pioneering, born digital HEI, defined as an organization where IT has played a central role since its conception, and whose growth has had a clear linkage to the use of digital technologies (Tumbas et al., 2015). Established in 1995 and headquartered in Spain, it is medium-sized, private but partially state-funded, with an international community of 4,000 remote professors. It has grown from 50,000 to 75,000 students in five years. It was the first university to operate exclusively online. It revolutionized higher education with its asynchronous online educational model and is considered a digital native. It is a global university born in the digital age that is willing to educate global and digitally skilled citizens, generating a positive social impact. Considered the world's first online university, it has a unique online methodology consisting in its proprietary learning model based on three elements: learning resources, personalized student support from teaching staff, and collaboration. Its 100% online methodology is unique, innovative, and internationally renowned.

The longitudinal approach is gained by the research being developed in two moments:

- Ex-ante (before the emergence of the pandemic; November 2018 to January 2019): In this stage, the first part of the interview guideline was adapted from a previous research work on BMI in Industry 4.0 (Müller et al., 2018) to include five blocks: (a) the interviewee profile; (b) the interviewee's understanding of the DT concept; (c) the DT process; (d) the tensions and solutions derived from DT for each of the BMI sub-constructs (Clauss, 2017), namely value creation, value proposition, and value capture; and, (e) the HEI's vision for the future due to the impact of DT. All the interviews were audio recorded and literally transcribed. The data were coded simultaneously but separately by two coders, who identified themes derived from the data with the aim of identifying meanings in the transcribed interviews (Corbin & Strauss, 2015). Sentences or groups of sentences were coded, compared (interrater agreement: 0.75), and discussed until agreement was reached on codification and analysis.

- Ex-post (after the emergence of the pandemic; July 2020-December 2020): In this stage, the investigation was structured around three temporal phases in relation to the pandemic: (i) COVID-19 emergency phase (March-June 2020), with topics including areas of the university most significantly impacted (teaching, research, transfer, others) due to forced digitalization, worse and better situations and how they were handled, and if the situation experienced impacted the vision of what DT is and its importance; (ii) New normality COVID-19 stage (July 2020-December 2020), with topics including the impact map of DT for the main HEI stakeholders (students, teaching and research staff, administration and services personal, companies, and society), main DT-derived challenges and how to overcome them, and main DT-derived opportunities and how to take advantage of them; (iii) Visions of the future, with the focus on understanding the HEI's vision for the next five years in the light of the impacts of DT and COVID-19. All the interviewees were administered a questionnaire via e-mail, and telephone support was provided where required.

Aside from the primary data gathered through interviews, information was provided by two of the authors who have had more than 25 years of combined experience in the HEI studied. The first collaborated from its foundation in 1995 until 2000, designing teaching materials and acting as a remote teacher. The second has been teaching remotely in the HEI since 1999, experiencing firsthand many of the digital transformation changes that have occurred over the last two decades. These two authors provided information via direct observation and access to internal and external communications through the intranet and the website, respectively.

4. Results

4.1. Digital transformation before and after the COVID-19 shock

4.1.1. External drivers of change

Increased collaboration, competitive pressure, and technology adoption: The results show that an expected strategy for overcoming this DT challenge is based on collaboration between HEIs themselves, governments, and industry, as stated by one of the participants: *"The challenges of technological change are so great that they push for collaboration [...]. It is mandatory for us to work together, otherwise we will not succeed."* (VRCE). This shock effect of the pandemic has had a catalytic effect on the institution, representing a turning point in its acceleration towards developing a new learning value proposition. Forced digitalization has suddenly created new competitors in both domestic and international markets. The vast majority of traditional HEIs have begun to develop online teaching in one of its multiple modalities, ranging from integrating video conferencing systems in the virtual campus to continue offering classes in synchronous mode to simply opening a discussion forum for questions.

Demand shock: The magnitude of the COVID-19 pandemic, its initial stage of total confinement, the subsequent new normality with its possibilities for face-to-face and online hybrid teaching, albeit under the enormous uncertainty of what will happen in the short-term future, have forced thousands of newcomers to the university world to consider educational options as no previous generation has. Eighteen-year-old, traditionally mostly face-to-face university students have suddenly become digital students, a target audience for whom the HEI studied is not prepared, having previously not been their focus, illustratively stated as: "*We say we are a complementary university to the university system because the face-to-face universities have already got the 18- to 25- year olds, whereas we have many of the rest of students*" (VRCE).

4.1.2. Effects on competition and digital transformation

Blurring competitive boundaries: In the pre-pandemic stage, the institution was already actively immersed in a continuous DT process: "*We are an online university, but we already needed this process of digital transformation, and we are now immersed in it."* (DMO). As a born digital HEI with a strong international presence, the institution had acknowledged the incipient entry of new competitors such as the technological giants in the world of higher education, and had already taken important steps to be able to prevail, including assigning a significant fixed annual budget for investment in technologies that would facilitate DT. Paradoxically, despite being born digital, the organization lacks the digital mentality: "*We do not know how to manage the efficiency that digital transformation can give us, and this is because we still do not have the digital mindset.*" (VRSPR). Consequently, the commitment of the management team and the governance and decision-making structures are perceived as necessary to overcome the different pockets of resistance.

In this pre-pandemic stage, the HEI saw the global digital technology companies (e.g., Google), EdTech unicorns (e.g., Udemy, Coursera), and start-ups as the only disruptive groups of competitors, aware that as a born digital university its value proposition was already clearly differentiated from traditional HEIs. Despite this incipient and growing threat, the priorities of digitalization are still closely linked with the search for efficiency and using digital technologies to do things better and save costs, while there is also increasing internal pressure to use DT to transform the *what* the HEI does. Pressure towards innovation is growing in the direction of personalizing the learning experience and offering learning programs and teaching methodologies focused on the development of the job market demanded skills rather than the simple issuance of official degrees. For example, one of the participants asked, *"Will we survive ourselves? We could die as a university because of not being able to offer this customization of the curriculum [...]. It will not be enough for companies if you have a master's degree ... their question is "Do you know how to solve this?"* (VRSPR). At the other extreme, the risk of digital fracture is also perceived by the students who do not follow the pace of online training.

The studied HEI will have to update the delivery medium via technology to be able to move towards offering the student a full online personalized learning experience, even if as a born digital player this delivery is instructor-mediated. e.g., offering virtual face-to-face synchronous sessions using videoconferences that will complement other asynchronous e-learning strategies such as discussion boards, e-mail, etc. Technology investments will be needed to allow customization to be scalable and automated, with artificial intelligence and data management included among the required technologies, and always with a mobile-first mentality. The studied HEI will have to make changes to the instruction type to move towards this personalised learning experience, with students receiving customized content based on the chosen curriculum and selected itinerary. Innovation is likewise required in how they apply content (e.g., "*Do we have to set everyone the same exam? Individualization and personalization of exams [...]."* (VRSPR*),* with different modes to deliver the activities, including video, audio, and text, based on personal preferences and type of learner.

Acceleration of DT: The COVID-19 shock has not had such a dramatic effect on the studied HEI as on the HE sector as a whole, at least in terms of online teacher education, which has been carried out digitally in the studied HEI for the last 25 years. However, it undoubtedly urgently increases the need to significantly accelerate the DT started, not only to redirect the situation in the short term (e.g., to work remotely), but above all to accelerate the strategic transformation towards a new value proposition in teaching, as illustrated in the following examples:

The COVID-19 effect has further impacted the need to fully implement digitalization. In recent years [...] much importance has been given to the transformation of the HEI, considered as an entity. The greatest impact has been the speed with which these changes have been made and the symbiosis that has been caused between the changes in the HEI and society itself, which has also advanced in a definitive way towards its digitalization (VRSPR).

4.2. Learning value proposition and business model changes

In reaction to the combined effect of emerging EdTechs and the forced digitalization of traditional HEIs, the born digital HEI feels pressured to accelerate the design of a new online learning value proposition that will act as a renewal engine, significantly impacting the different dimensions of the current business model (Table 2), namely value proposition, value communication, value creation, value delivery, and value capture (Abdelkafi et al., 2013). As regards the value proposition, the HEI is clearly aiming towards a more student-centric lifelong learning relationship model, a crucial aspect of which is offering the student a digital experience (SX) at the level of the best practices of global benchmarks. The strategy to achieve this SX is clear: enabling a new personalized online value proposition for each student and becoming a guide for the student before (what to study?), during (how to improve teaching?), and after finishing a particular program (how to improve employability?). This new vision impacts the entire learning value proposition, not only changing what the HEI wants to offer (e.g., adding new short-term professionalizing programs based on skills development, offering MOOCs), but also the typology of teaching materials (e.g., more multimedia materials, curated from third parties) and how they are distributed (based on personalized curriculums and itineraries, recommendations, etc.) and consumed (interactively with the professor, with a flexible self-paced approach). A fundamental aspect of this new learning value proposition is the significant increase in the number of different learning methodologies and activities offered, creating a digital ecosystem of multimode learning methods and tools. These include, among others, direct instruction via synchronous video conferences and asynchronous videos, project-based learning, employer-based learning, mobile learning, peer-to-peer learning, simulators, self-assessment tests, online exams with identification of the person, and authorship of the content. Included as important additional benefits of this renewed learning value proposition are a new student-trainer relationship supported not only by multimode learning tools but also by artificial intelligence and data analytics, and access to a customized virtual campus developed with a mobile-first mentality, highlighting the need to deploy both digital and educational technologies. As stated by different participants: *"Regarding the offering, the real opportunity is the idea of being able to offer personalization [...], such as enabling students to decide their own curriculum. Some students are already asking for this and we are not able to offer it."* (VRSPR); "*Learning resources end up being much more multimedia [...] There is text, there is video, there is audio, there are other types of resources such as simulators..."* (VPOT); *"More customizable teaching, and we can customize itineraries. Here we have challenges that without the new technologies we would not even consider."* (VRCE).

Personalization means one by one, therefore you should be able to progress at the student's pace; and while this is true for teaching it is still lacking for assessment and examination [...]. We have now achieved monthly enrolment but imagine there were 365 different enrolment periods, every day of the year and whenever the student wants [...]. It means a different organization [...]. This is not feasible without artificial intelligence … because otherwise the question is, what is the alternative? Having as many teachers as students? (VRSPR).

As regards **value creation**, the HEI will need to put the appropriate combination of own resources and activities and those contributed by partners to work to create a new learning value proposition that leverages both digital and education technologies and capabilities. Undoubtedly, a key resource is and will continue to be the virtual campus as the motor for configuring the personalization of the learning experience and customizing curricula, itineraries, and paces of study, providing access to a multitude of types of both received and applied content and tools for maximising student-professor interaction. New skills and mindsets are needed to be able to create this value, such as detailed planning of all teaching activity: *"You need to plan everything carefully, there is no window for improvisation."* (VRSPR); a more open concept in terms of technology, for example a "*Lego style platform.*" (VRCE), enabling third party technologies and capabilities that incorporate artificial intelligence and data analytics to be "plugged in": "*It is teaching improvement based on data analytics, not so much intuitively [...] but systematically monitoring what happens in classrooms, and we do that through technology*." (VRCE); new operative processes, for example enrolment 365 days a year; and technologies to guarantee the identification of the student and authorship of the content of exams, among many others.

As regards **value delivery**, the HEI will have to update the customer segments targeted, the distribution channels used, and the customer relationship developed to deliver this new learning value proposition. Getting to know the students better is crucial to be able to offer tailored automated learning services (contents, methodologies, assessments, etc.) in a co-creation environment and with individual support. Ubiquity due to the mobile phone, "*The University is in your pocket, in space and in tim*e." (VRCE), and social networks will be fundamental to interact with students as the prior importance of the teaching classroom decreases, as stated by one participant: "*A challenge is that the classrooms disappear [...]. We are in the digital world [...] but in fact we continue thinking about students and classrooms.*" (VRSPR). To implement this value delivery mechanism, the appropriate amount and combinations of digital educational technologies will need to be deployed.

As regards **value capture,** this new learning value proposition will represent not only new sources of revenue (e.g., shorter professionalizing programs) and new forms of revenue (e.g., subscription-based), but also new challenges in the cost structure. These include continuous investment in technology (many of them digital educational technologies), attracting digital talent, software licences, cloud services, and the creation of content offered for free (e.g., MOOCs). An illustrative statement is: *"Will subscription happen with university services? Services that you subscribe to, and depending on the level of subscription you have, you will be able -or not- to access a typology of course.''* (VRCE).

As regards **value communication,** this new learning value proposition must be translated into a highly attractive storytelling narrative that connects in both a relevant (narrowcasting approach) and an automatic way:

[...] better and more personalization of the student experience and maybe what we offer them and our relationship with them. Therefore, there is a great opportunity for us to know more about the student and prepare a valuable customer journey from awareness of our offering to employability (VPOT).

4.3. Roadmap towards a customized multimode learning strategy

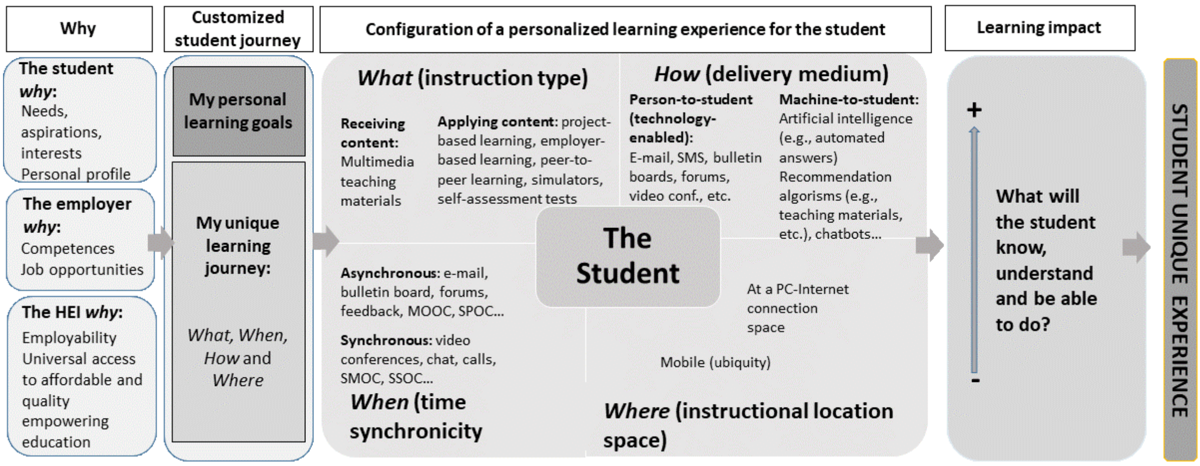
In response to these anticipated changes at the level of the different dimensions of the business model, the HEI is designing its roadmap towards a customized multimode learning strategy that will change the *what, when, how,* and *where* of the learning journey (Figure 1). In this roadmap, the HEI helps to match students' needs, aspirations, and interests with opportunities (e.g., professional-related, discovery-related), which is the basis for establishing the customized student learning briefing (learning objectives and desired learning journey). Once their goals are established, the student gets automatic recommendations and can configure a personalized learning experience that covers (a) the what (instruction type), both for receiving content (multimedia teaching materials) and applying content (project-based learning, employer-based learning, peer-to-peer learning, simulators, self-assessment tests, etc.); (b) the how (delivery medium), including person-to-student (technology-enabled; e-mail, SMS, bulletin boards, forums, video conferences, etc.), machine-to-student (artificial intelligence such as automated answers), and recommendation algorithms (e.g., teaching materials), chatbots…; (c) the when (time synchronicity), both asynchronously (e-mail, bulletin board, forums, feedback, MOOC, SPOC, etc.) and synchronously (video conferences, chat, calls, SMOC, SSOC, etc.); and (d) the where (instructional location space), both in a PC-Internet connection space or on a mobile device. This "mobile-first mentality" in the development of digital technologies will lead to a ubiquitous learning mode, making learning possible any time and in any place. Following this individually configured online multimode learning journey, the student will undergo a "learning impact" (what the student will know, understand and be able to do) and a unique student experience (SX).

Table 2. Envisioned business model of the digitalised university

Value communication	**Value proposition**	**Value creation**
Offer and promote the best learning experience to the 14-18-year-old segment; Use CRM software and digital communication tools and channels to automate demand management and campaigns; Use social media to create and	Role of the HEI Offer the best global student digital experience (SX) before deciding what to study, during the learning process, and afterwards (relational); Personalize learning: tailoring learning for each student based on needs, interests, aspirations, and background; Customize the what, how, when, and where the students learn; Guide the student throughout the learning journey; Provide the (potential) student free content (MOOC's); Facilitate a self-managed digital student in the job market; Build a community of innovation, creativity, and entrepreneurship; Promote sustainability and break the digital barrier.	Design a customized virtual campus with a variety of learning methodologies to deliver a fully personalized student journey with high flexibility (self-paced); A more open concept in terms of technology ("Plug and Play"); Mobile-first mentality; Create the best multimedia teaching materials available online; Alliances and active role in

communicate attractive targeted promo materials and campaigns.	<u>Selection, distribution, and consumption of contents</u> Personalization of curriculums and customization of learning itineraries; New courses and contents based on a faster connection to the labour market (new offering); Access to a learning resource hub with more contents and multimedia resources; multimedia resource centre that integrates both proprietary teaching materials and those developed by third parties; Value-added interactive teaching materials, with marks, comments, etc. to support the student. <u>Design of learning methodologies and activities</u> New forms of applying content; dual training, professional final projects, simulators, etc.; High flexibility, self-paced learning approach; Blending a variety of learning online modes and methods: direct instruction via synchronous video conferences and asynchronous videos, project-based learning, employer-based learning, mobile learning, peer-to-peer learning, simulators, self-assessment tests, etc.; New forms of virtual internships. <u>Assessment</u> Online exams with identification of the person and authorship of the content; Certified guarantee of identification of the student and authorship of the content. <u>Student – Professor interactions</u> Synchronous video conference interactions, both individual and group; Asynchronous video interactions, both for one individual or a group; Improved teaching process by incorporating data analytics; Improved teaching process complemented with artificial intelligence; Choose or being assigned the best expert based on the student's teaching needs; Getting to know the students better to offer tailored automated learning services (contents, methodologies, assessments, etc.) in a co-creation environment and with individual support. <u>Virtual campus and Technology strategy</u> A customized virtual campus with a variety of learning methodologies to deliver a full personalized student journey with high flexibility (self-paced); Total mobile ubiquity.	the wider eco-system; Online exams with identification of the person and authorship of the content; Artificial intelligence and analytical data to systematically improve teaching; SasS subscription-based payments; Micro-monitor the competences developed by each student; Scalability due to digitalization (e.g., student support); e.g., customization based on industrialization and scalability.
Value capture Subscription; New sources of revenue from new "professionals", shorter programs; Receiving content that is free (e.g., MOOCs); Re-invest cost savings in added value for the student; Attract digital talent (data, analytics, cybersecurity, etc.); Fixed investment and continuous renewal of technology.		**Value delivery** One by one interaction (online classroom disappears); Ubiquity thanks to the mobile; A digital licensing system (digital teaching materials developed by third parties); The student sets the pace of study and examinations (time is variable); New social media support channels; The new channels to connect fast with the current and new markets; Getting to know the students better to offer tailored automated learning services; Getting to know the students better to offer tailored automated learning services (contents, methodologies, assessments, etc.) in a co-creation environment and with individual support.

Figure 1. Roadmap for a born digital HEI towards a customized multimode learning strategy



To be able to deliver this customized multimode learning strategy, the HEI will deploy new digital and educational technologies and capabilities that will impact the different business model dimensions (Table 2), including: (i) social CRM software (to be a guide for students prior to enrolling, during the learning process, and after graduation); (ii) profiling and customization software (to personalize curriculums, assessments, self-pace, etc.); (iii) portfolios of online learning modes and methods (e.g., direct instruction via synchronous video conferences and asynchronous videos, project-based learning, employer-based learning, mobile learning, peer-to-peer learning, simulators, self-assessment tests, etc.); (iv) the resource platform (to integrate third party multimedia resources) and interactive teaching resources (with marks, comments, etc.); (v) recommendation engines (e.g., library); (vi) student identification software (e.g., for online assessments); (vii) authorship software (e.g., to avoid plagiarism); (viii) data analytics (e.g., to learn better teaching practices); (ix) artificial intelligence (e.g., to support the professor); (x) mobile-first mentality, technology integration, and partnerships (e.g., Google Workspace for Education Fundamentals); and (xi) 365 days a year enrolling software (e.g., the concept of classroom disappears).

5. Discussion

5.1. Contribution of the paper

In the pre-pandemic stage, our results on DT as a necessary continuous process, including for a born digital HEI, confirm previous research (Zulfikar et al., 2018; Khalid et al., 2018). The findings on the need to offer a ubiquitous learning mode through a mobile device respond to research calls (Kinshuk et al., 2013) and confirm how digital technologies are becoming inevitable (Albirini, 2007) and are disrupting universities (Posselt et al., 2018), and especially but not exclusively the traditional HEIs. Research has also confirmed how new digital asynchronous and synchronous applications are changing the learning process, placing insufficiently adapted HEIs at risk of irrelevance (Kaplan & Haenlein, 2016), as happened at least temporally to almost a third of HEIs during the very first stage of the COVID-19 forced digitalization (Marinoni et al., 2020). This paper contributes further empirical evidence by showing that even though it was born digital the organization lacks a digital mentality, in line with previous research that points to the importance of addressing the academic digital gap during DT processes (Bond et al., 2018). The results also evidence that a shock such as COVID-19 is a cultural change that can eliminate digital resistances practically immediately, accelerating the digitalization mentality and processes by means of working/studying remotely and online exams. This confirms recent extant research on how the forced learning and testing of new digital tools and methodologies experienced by teachers during the pandemic has changed their digital mentality (Marinoni et al., 2020), even when new skills and mindsets such as planning all teaching activity in great detail are needed (Nguyen, 2018). The results also show that the magnitude of the COVID-19 exogenous shock (Fligstein & McAdam, 2011, p. 32) for the entire higher education sector has shown that HEIs need to be more business oriented to survive, contributing further empirical evidence that business practices are becoming a necessity in HEIs (Pucciarelli & Kaplan, 2016). Regarding the impact of DT on the BM, our results contribute to some recent attempts to connect the business model concept with the field of universities (Abdelkafi et al., 2018; Rosi et al., 2018; Posselt et al., 2018; Rof et al., 2020).

One of the primary effects of the COVID-19 shock is an acceleration of the HEI intention to design a personalized online value proposition (customer-centric). This finding is consistent with previous research showing how adding digital technologies can contribute to developing new interactive and co-creation-based learning strategies (Pucciarelli & Kaplan, 2016), and how e-learning is more learner-centric and learner-personalized, supported by the always changing digital technologies that offer ubiquity in the access and delivery of teaching resources and services anytime, anywhere, in an interactive, collaborative and personalized manner (Nguyen et al., 2019). The COVID-19 pandemic has not brought about a technological jump since digital technologies were previously available and disrupting the sector (Posselt et al., 2018), but rather it has generated a cultural jump that has caused a new digital mind-set (Marinoni et al., 2020), removing, or at least making inoperative, resistance to change. The state of emergency has automatically answered the key questions "*Is this the moment*", "*Is it really necessary?*" "*Are we ready?*" and "*Is this the solution?*" in the affirmative, facilitating the adoption of new technologies and learning systems (Agasisti et al., 2020; Izagirre-Olaizola & Morandeira-Arca, 2020; El-Bassiouny & Mohamed, 2020; Dietrich et al., 2020). The finding about the importance of creating a digital ecosystem of multimode learning methods and tools (e.g., direct instruction via synchronous video conferences and asynchronous videos, project-based learning) for this learning value proposition is in line with previous research on different forms of instruction, explaining online distance learning (Kaplan & Haenlein, 2016). The results at the level of the complementary BM dimensions that contribute to creating this new learning value proposition (value creation), delivering it (value delivery), generating new sources of revenue and costs associated with it (value capture), and the way to communicate it (value communication), show the necessary interconnection between the different building blocks of the business model (Osterwalder et al., 2010; Abdelkafi et al., 2013). The findings also clearly show that to innovate the BM several of its dimensions must be changed simultaneously (Winter & Szulanski, 2001; Johnson et al., 2008; Baden-Fuller & Haefliger, 2013; Baden-Fuller & Mangematin, 2013).

This article contributes to the previous debate on learning theory associated with educational technologies (Albirini, 2007) and responds to recent calls for further personalized learning research (Xie et al., 2019). Although most of the empirical results obtained point to the development of a customized multimode learning strategy that shares the basic principles of the constructivist theory, the reality is that, in its pure state, the constructivist theory can generate certain problems, especially regarding "knowledge construction" (Gärdenfors & Johansson, 2005). There are different ways to meet students' unique learning needs and at least two will use technology and multimodality: (i) customization, leaving the agency (the choice of multimodality options) to students; and (ii) personalization (using data and algorithms to create a personalized learning by leveraging multimodality options). In the latter case, new technologies that decide for the student, such as learning analytics and artificial intelligence, can open the pathway to methodologies that are closer to behaviourism through personalizing learning journeys for students with similar learning profiles (e.g., adaptive learning technology) in a scalable manner. The optimal learning paths are likely to be somewhere between the two strategies, combining the best of customization (constructivism) and personalization (likely behaviourism), thus contributing to satisfying a diversity of students' cognitive styles (Hsieh & Chen, 2016).

5.2. Managerial implications for HEIs

This empirical research has several practical implications. The findings presented provide "out-of-the-box" tools and frameworks that can encourage reflection, help design a student-centric multimodal learning value proposition, and facilitate the required changes to the BM. The analysis is of great value for the entire higher education sector, including both born digital and traditional HEIs, because as the competitive boundaries blur due to digitalization participants become potential international competitors of all the others.

HEI managers could use the "Envisioned Business Model of the Digitalised University" framework (as exemplified on Table 2) to benchmark with the innovative EdTech to find sources of differentiation, and to prioritize decisions and plans about building and managing the right digital and educational technologies ecosystem (e.g., direct instruction via synchronous video conferences and asynchronous videos, group-project-based learning, online exams, etc.). This framework, as a practical tool for strategic reflection, could also be used to explore the trade-offs between the concepts of cost-efficiency, effective education, and continuous innovation, a topic that calls for further investigation (Stepanyan et al., 2013). It could also be used internally (employees) and externally (students and other relevant stakeholders) to test ideas, design new ideas (e.g., in a participative way to build shared vision), and communicate results.

Second, HEI managers could reflect and build their "Envisioned Business Model of the Digitalised University" to create an overview of the desired business model associated with this new multimode learning value proposition, and to deploy the required digital and educational technologies. Detailed specifics of the BM dimensions would enable DT, academic, and organizational "going toward" plans to be formulated: a) At the level of learning value proposition: clarification and reflection on the role of the HEI, the selection, distribution, and consumption of content, the design of learning methodologies and activities, assessments, student-professor interactions, and virtual campus and technology strategy; b) At the level of value creation: resources, activities, and partnerships to create this new learning value proposition; c) At the level of the value delivery: customer segments targeted, the distribution channels used, and the customer relationship developed to deliver this new learning value proposition; d) At the level of value capture: sources of revenue and cost structure associated with this new learning value proposition; and e) At the level of value communication: how this new learning value proposition will be translated in a highly attractive storytelling narrative that connects in both a relevant and automatic way.

Third, HEI managers could use the "Roadmap for a born digital HEI towards a customized multimode learning strategy" framework (Figure 1) to visualize the desired student-centric learning strategy. This tool would also be useful for internal communication, further driving opportunities to develop an interactive version to communicate the value proposition to the potential student community.

6. Conclusion

The COVID-19 pandemic has shaken up the entire higher education system, causing a forced and severe shift in the scale of DT, which became mandatory to remain operational during the shock, moving from a situation of "*an ongoing digitalization process*" to a situation of "digitalize now or stop operating." It can be argued that the essence of the COVID-19 effect has been more of a "real-time" cultural transformation than a DT one, at least for two thirds of HEIs. Before the pandemic, digital technologies were there to be used. EdTech players had already detected this opportunity, but resistance forces were at play in the more traditional HEIs. However, the outbreak of the pandemic and especially the lockdown meant the immediate elimination of all resistance.

In this context, our study responds to the call for more research on the impact of COVID-19 in the HEI sector, empirically exploring the case of a born digital HEI and providing an analysis of the changes that have taken place since the COVID-19 shock. This paper contributes to the limited literature on the learning value proposition of HEIs as the core component of their BM, but within a more global and interdependent HEI BM. The business approach to HEIs allows for a better analysis of their requirements for competitiveness and survival as organizations in a competitive sector. Second, the analyses made describe the decision and visions both prior to and post the emergence of the COVID-19 pandemic, uncovering the practice of digital transformation and how it has been accelerated by the shock. The findings and discussion uncover the sources of organizational challenges for HEIs (managers, teachers, and staff) in their digital transformation. Third, the importance, nature, and possible evolution of learning multimodality is described and analysed in this DT context. And fourth, this research contributes by designing a roadmap towards this customized multimode learning strategy to offer a unique personalized learning journey for each student based on goals, preferences, and cognitive styles (Hsieh & Chen, 2016), among others. In a global sense, this research provides empirical evidence and is a critical analysis at the intersection of the HEI business model's digital transformation in response to the COVID-19 shock.

This paper is subject to some limitations regarding its methodology and findings. The contribution is limited due to the use of a single case study from a specific sector, so it should be considered exploratory and theory-grounding research. Future research should validate our findings and respond to some unanswered questions, the first of which is whether the effect of COVID-19 forced the need for multimodality and personalization. This paper argues that this was a forced test and that higher education will be transformed to deliver personalized multimode learning value propositions. This personalization will require decisions about technology models (Alamri et al., 2021), the development of a variety of technological tools aligned with different ways to learn (Stepanyan et al., 2013), and a general cultural shift (Renz & Hilbig, 2020). The business model vision will be fully integrated into HEIs' decision-making and management processes. What is not clear is whether and how this multimodality will be used for differentiation among HEIs and other education suppliers, allowing for different types of learning value propositions, or whether students will demand the maximum customization of all education offers and all suppliers will evolve towards the same standards of multimodal customization. However, the degree to which the COVID-19 state of mind and practice as regards digitalization and customization has been implemented during the first year of the pandemic has been at a huge and unsustainable cost to HEIs and their staff. Thus, questions arise about the degree to which this forced digitalization will have a permanent cultural effect or will it be eroded when the situation goes back to "normal" or stays stable in a "new normality" scenario: *Is it a lost war for some of the stakeholders? For example, for teachers required to be available 24X7?* We wonder whether this digitally prone mindset will continue among HEI managers, teachers, and staff so that current methods cease to be used and the new emerging ones fully adopted. Any forced organizational change may be subject to possible setbacks and restraining forces (Alvesson & Sveningsson, 2015).

Any relevant level of customization or personalization faces the problem of scalability in the sense of being able to personalize the learning experience for many students, including international students with a high cultural diversity, making it necessary to offer different options in terms of technology-assisted learning tools (Habib et al., 2014). This will require investment in both digital and educational technology to allow for automation, creating a technological challenge for the delivery medium (Margulieux et al., 2016), which can be resolved using artificial intelligence applications (Renz & Hilbig, 2020). Nonetheless, this is likely to pose important challenges for the management team, raising the question, Will HEIs become like EdTech players? Aside from managerial and other organizational barriers to the adoption of artificial intelligence solutions (Renz & Hilbig, 2020), more research and experimentation is needed to test whether promises made to produce satisfaction on each personalized learning journey are kept, especially given that a cultural change is needed (Renz & Hilbig, 2020). Where these technologies are used successfully it will be interesting to further explore how they will combine with real-people (teachers, tutors, staff) support and how this will change the role of teaching and non-teaching HEI staff. Further research must also be developed on the impact of artificial intelligence on the BM and the return on investment (Stepanyan et al., 2013).

Inequality concerns are another social challenge for the DT of HEIs. According to (Marinoni et al., 2020), a third of HEIs did not adapt fast enough to the new digitalization forced by COVID-19, begging the question as to how many HEI students have consequently been unable to catch up. There is also the risk of digital fracture for students who do not follow the pace of online education, as has already been shown in studies that suggest that the degree of suitability of web-based lectures depends on the characteristics of the student (Montrieux et al., 2015), being less suitable for low achieving students (Owston et al., 2013). This raises the issue of what the HEI will offer these students and will they be able to deal with this problem, or alternatively will it become a social one? Although university students are generally highly skilled for technology adoption, some technologies may require more sophisticated infrastructure and ICT competences, which might not be available or evenly distributed among students in different geographical areas and with varying economic statuses. These challenges may be insurmountable for HEIs and need a systemic public approach. In this line, collaboration among HEIs, the government, and even industry may be necessary for a smart and inclusive DT of higher education.

Our paper points to a highly customized unique student experience delivered in a multimode learning modality, further questioning how quality is perceived and predictably understood, valued, and interpreted in a way in which traditional quality becomes obsolete and excellence and delight gain prominence. It is relevant to know what students value in terms of learning/training quality, platform quality, study material quality, and learning experience quality, to mention just a few, as they seek human interaction in their learning path, conditioning the degree and quality of Artificial Intelligence applications in the HEI sector (Renz & Hilbig, 2020). Regarding the issue of quality, several questions can be asked from a behaviourist point of view. For example, will students be capable of constructing their learning packs or paths (e.g., when choosing the open digital badge or the competency-based learning program)? Will learning be constructed in the right way (effectively, efficiently, etc.)? And from a strategic point of view, we may ask what model of personalization will universities adopt in the future, how the collective intelligence of experienced professors will be leveraged, who will lead this future customized multimode learning strategy, the student, the professor, or the algorithms, and how will these decisions affect quality?

Last, our study shows the ingredients needed for *technology acceptance, questioning the diffusion and perdurance of the outcoming innovation*. Further research could tackle already traditional approaches in the field of innovation (Technology Acceptance Model and the Diffusion of Innovation Theory) and test their robustness and universality in new and critical circumstances. Some experts predict that COVID-19 is just a first materialization of a series of shocks that will intensify and become more frequent due to climate change and derivates. It appears that we need to prepare organizations and future generations to cope with these shocks and manage transformation processes in a sustainable way, and HEIs and the public sector serve as an appropriate illustrative example.

Acknowledgement

This work was supported by the Ministerio de Economía, Industria y Competitividad/Agencia Española de Investigación/ FEDER/ UE [grant number ECO2017-86054-C3-3-R].

References

Abdelkafi, N., Hilbig, R., & Laudien, S. M. (2018). Business models of entrepreneurial universities in the area of vocational education - An Exploratory analysis. *International Journal of Technology Management, 77*(1/2/3), 86-108. https://www.inderscienceonline.com/doi/abs/10.1504/IJTM.2018.091716

Abdelkafi, N., Makhotin, S., & Posselt, T. (2013). Business model innovations for electric mobility-what can be learned from existing business model patterns? *International Journal of Innovation Management, 17*(1). https://doi.org/10.1142/S1363919613400033

Agasisti, T., Frattini, F., & Soncin, M. (2020). Digital innovation in times of emergency: Reactions from a school of management in Italy. *Sustainability (Switzerland), 12*(24), 1–17. https://doi.org/10.3390/su122410312

Alamri, H. A., Watson, S., & Watson, W. (2021). Learning technology models that support personalization within blended learning environments in higher education. *TechTrends, 65*(1), 62–78. https://doi.org/10.1007/s11528-020-00530-3

Albirini, A. (2007). The Crisis of educational technology, and the prospect of reinventing education. *Educational Technology & Society, 10*(1), 227–236.

Alvesson, M., & Sveningsson, S. (2015). *Changing organizational culture: Cultural change work in progress*. Routledge.

Baden-Fuller, C., & Haefliger, S. (2013). Business models and technological innovation. *Long Range Planning, 46*(6), 419–426. https://doi.org/10.1016/j.lrp.2013.08.023

Baden-Fuller, C., & Mangematin, V. (2013). Business models: A Challenging agenda. *Strategic Organization, 11*(4), 418–427. https://doi.org/10.1177/1476127013510112

Bhardwaj, A.K., Garg, L., Garg, A., & Gajpal, Y. (2021). E-Learning during COVID-19 outbreak: Cloud computing adoption in indian public universities. *Computers, Materials & Continua, 66*(3), 2471–2492. https://www.techscience.com/cmc/v66n3/41083

Bond, M., Marín, V. I., Dolch, C., Bedenlier, S., & Zawacki-Richter, O. (2018). Digital transformation in German higher education: student and teacher perceptions and usage of digital media. *International Journal of Educational Technology in Higher Education, 15*(1), 48. https://doi.org/10.1186/s41239-018-0130-1

Clark, R. & Salomon, G. (1986). Why should we expect media to teach anyone anything. In R. Clark (Ed.), *Learning from Media: arguments, analysis, and evidence* (pp. 37-70). Information Age Publishing.

Clauss, T. (2017). Measuring business model innovation: conceptualization, scale development, and proof of performance. *R&D Management, 47*(3), 385–403. https://doi.org/10.1111/radm.12186

Corbin, J., & Strauss, A. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory.* (4th ed.). Sage Publications Inc. https://us.sagepub.com/en-us/nam/basics-of-qualitative-research/book235578

Corbo, L., Pirolo, L., & Rodrigues, V. (2018). Business model adaptation in response to an exogenous shock: An Empirical analysis of the Portuguese footwear industry. *International Journal of Engineering Business Management, 10*. https://doi.org/10.1177/1847979018772742

Demil, B., Lecocq, X., Ricart, J. E., & Zott, C. (2015). Introduction to the SEJ special issue on business models: Business models within the domain of strategic entrepreneurship. *Strategic Entrepreneurship Journal, 9*(1), 1–11. https://doi.org/10.1002/sej.1194

Dietrich, N., Kentheswaran, K., Ahmadi, A., Teychené, J., Bessière, Y., Alfenore, S., Laborie, S., Bastoul, D., Loubière, K., Guigui, C., Sperandio, M., Barna, L., Paul, E., Cabassud, C., Liné, A., & Hébrard, G. (2020). Attempts, successes, and failures of distance learning in the time of COVID-19. *Journal of Chemical Education, 97*(9), 2448–2457. https://doi.org/10.1021/acs.jchemed.0c00717

El-Bassiouny, N., & Mohamed, E. K. A. (2020). Replacing exams with research papers: Chronicles of a higher education institution (HEI) amidst COVID-19 pandemic. *Journal of Islamic Marketing, 12*(3), 627-630. https://doi.org/10.1108/JIMA-08-2020-0226

Felder, R., & Silverman, L. (1988). Learning and teaching styles in engineering education. *Engineering Education, 78*(7), 674–681.

Findlater, L., & McGrenere, J. (2004). A Comparison of static, adaptive, and adaptable menus. *Conference on Human Factors in Computing Systems - Proceedings* (pp. 89–96). https://doi.org/10.1145/985692.985704

Fligstein, N., & McAdam, D. (2011). Toward a general theory of strategic action fields. *Sociological Theory, 29*(1), 1–26. https://doi.org/10.1111/j.1467-9558.2010.01385.x

Foss, N. J., & Saebi, T. (2017). Fifteen years of research on business model innovation. *Journal of Management, 43*(1), 200–227. https://doi.org/10.1177/0149206316675927

Frias-Martinez, E., Chen, S. Y., & Liu, X. (2009). Evaluation of a personalized digital library based on cognitive styles: Adaptivity vs. adaptability. *International Journal of Information Management, 29*(1), 48–56. https://doi.org/10.1016/j.ijinfomgt.2008.01.012

Gärdenfors, P., & Johansson, P. (2005). Cognition education and communication technology. In *Cognition, Education, and Communication Technology* (pp. 1-294). Lawrence Erlbaum Associates.

Gong, C., & Ribiere, V. (2021). Developing a unified definition of digital transformation. *Technovation, 102*, 102217. https://doi.org/10.1016/j.technovation.2020.102217

Habib, L., Johannesen, M., & Øgrim, L. (2014). Experiences and challenges of international students in technology-rich learning environments. *Educational Technology & Society, 17*(2), 196–206.

Hsieh, C. W., & Chen, S. Y. (2016). A Cognitive style perspective to handheld devices: Customization vs. personalization. *International Review of Research in Open and Distance Learning, 17*(1), 1–22. https://doi.org/10.19173/irrodl.v17i1.2168

Izagirre-Olaizola, J., & Morandeira-Arca, J. (2020). Business management teaching–learning processes in times of pandemic: Flipped classroom at a distance. *Sustainability (Switzerland), 12*(23), 1–18. https://doi.org/10.3390/su122310137

Johnson, M. W., Christensen, C. M., & Kagermann, H. (2008). Reinventing your business model. *Harvard Business Review, 86*(12), 57-68.

Kaplan, A. (2020). *Universities, be aware: Start-ups strip away your glory*. EFMD Global blog. https://blog.efmdglobal.org/2020/05/11/universities-be-aware-start-ups-strip-away-your-glory/

Kaplan, A., & Haenlein, M. (2016). Higher education and the digital revolution: About MOOCs, SPOCs, social media, and the Cookie Monster. *Business Horizons, 59*(4), 441–450. https://doi.org/10.1016/j.bushor.2016.03.008

Kay, J. (2001). Learner control. *User Modeling and User-Adapted Interaction, 11*(1), 111–127. https://doi.org/10.1023/A:1011194803800

Khalid, J., Ram, B. R., Soliman, M., Ali, A. J., Khaleel, M., & Islam, M. S. (2018). Promising digital university: A Pivotal need for higher education transformation. *International Journal of Management in Education, 12*(3), 264-275. https://www.inderscienceonline.com/doi/abs/10.1504/IJMIE.2018.092868

Kinshuk, Huang, H. W., Sampson, D., & Chen, N. S. (2013). Trends in educational technology through the lens of the highly cited articles published in the journal of Educational Technology and Society. *Educational Technology & Society, 16*(2), 3–20.

Kobsa, A., Koenemann, J., & Pohl, W. (2001). Personalised hypermedia presentation techniques for improving online customer relationships. *Knowledge Engineering Review, 16*(2), 111–155. https://doi.org/10.1017/S0269888901000108

Kuckertz, A., Brändle, L., Gaudig, A., Hinderer, S., Morales Reyes, C. A., Prochotta, A., Steinbrink, K. M., & Berger, E. S. C. (2020). Startups in times of crisis – A Rapid response to the COVID-19 pandemic. *Journal of Business Venturing Insights, 13*, e00169. https://doi.org/10.1016/j.jbvi.2020.e00169

Lee, D., Huh, Y., Lin, C. Y., & Reigeluth, C. M. (2018). Technology functions for personalized learning in learner-centered schools. *Educational Technology Research and Development, 66*(5), 1269-1302. https://doi.org/10.1007/s11423-018-9615-9

Magretta, J. (2002). Why business models matter. *Harvard Business Review*, *80*(5), 86–87.

Margulieux, L. E., McCracken, W. M., & Catrambone, R. (2016). A Taxonomy to define courses that mix face-to-face and online learning. In *Educational Research Review* (Vol. 19, pp. 104–118). Elsevier Ltd. https://doi.org/10.1016/j.edurev.2016.07.001

Marinoni, G., van't Land, H., & Trine Jensen, T. (2020). *COVID-19: Higher Education challenges and responses*. IAU. https://www.iau-aiu.net/COVID-19-Higher-Education-challenges-and-responses

Montrieux, H., Vangestel, S., Raes, A., Matthys, P., & Schellens, T. (2015). Blending face-to-face higher education with web-based lectures: Comparing different didactical application scenarios. *Educational Technology & Society, 18*(1), 170–182.

Morgan, T., Anokhin, S., Ofstein, L., & Friske, W. (2020). SME response to major exogenous shocks: The Bright and dark sides of business model pivoting. *International Small Business Journal: Researching Entrepreneurship, 38*(5), 369–379. https://doi.org/10.1177/0266242620936590

Müller, J. M., Buliga, O., & Voigt, K.-I. (2018). Fortune favors the prepared: How SMEs approach business model innovations in Industry 4.0. *Technological Forecasting and Social Change, 132*, 2–17. https://doi.org/10.1016/j.techfore.2017.12.019

Nandy, M., Lodh, S., & Tang, A. (2021). Lessons from COVID-19 and a resilience model for higher education. *Industry and Higher Education, 35*(1), 3-9. https://doi.org/10.1177/0950422220962696

Nguyen, D. (2018). The University in a world of digital technologies: Tensions and challenges. *Australasian Marketing Journal (AMJ), 26*(2), 79–82. https://doi.org/10.1016/j.ausmj.2018.05.012

Nguyen, Q. L. H. T. T., Nguyen, P., & Huynh, V. D. B. (2019). Roles of e-learning in higher education. *Journal of Critical Reviews, 6*(4), 7–13.

Osterwalder, A., Pigneur, Y., Clark, T., & Smith, A. (2010). *Business model generation: A Handbook for visionaries, game changers, and challengers*. John Wiley & Sons Inc.

Owston, R., York, D., & Murtha, S. (2013). Student perceptions and achievement in a university blended learning strategic initiative. *Internet and Higher Education, 18*, 38–46. https://doi.org/10.1016/j.iheduc.2012.12.003

Posselt, T., Abdelkafi, N., Fischer, L., & Tangour, C. (2018). Opportunities and challenges of Higher Education institutions in Europe: An analysis from a business model perspective. *Higher Education Quarterly, 73*(1), 100-115. https://doi.org/10.1111/hequ.12192

Pucciarelli, F., & Kaplan, A. (2016). Competition and strategy in higher education: Managing complexity and uncertainty. *Business Horizons, 59*(3), 311–320. https://doi.org/10.1016/j.bushor.2016.01.003

Renz, A., & Hilbig, R. (2020). Prerequisites for artificial intelligence in further education: identification of drivers, barriers, and business models of educational technology companies. *International Journal of Educational Technology in Higher Education, 17*(1), 14. https://doi.org/10.1186/s41239-020-00193-3

Rof, A., Bikfalvi, A., & Marques, P. (2020). Digital transformation for business model innovation in higher education: Overcoming the tensions. *Sustainability, 12*(12), 4980. https://doi.org/10.3390/su12124980

Rosi, M., Tuček, D., Potočan, V., & Jurše, M. (2018). Market orientation of business schools: a development opportunity for the business model of University Business Schools in transition countries. *E+M Ekonomie a Management, 21*(4), 175–194. http://hdl.handle.net/11025/31104

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. https://doi.org/10.1016/j.compedu.2019.03.006

Sejzi, A. A., & Aris, B. bin. (2012). Constructivist approach in virtual universities. *Procedia - Social and Behavioral Sciences, 56*, 426–431. https://doi.org/10.1016/j.sbspro.2012.09.672

Shemshack, A., & Spector, J. M. (2020). A Systematic literature review of personalized learning terms. *Smart Learning Environments, 7*(1). https://doi.org/10.1186/s40561-020-00140-9

Soluk, J., Kammerlander, N., & De Massis, A. (2021). Exogenous shocks and the adaptive capacity of family firms: exploring behavioral changes and digital technologies in the COVID-19 pandemic. *R&D Management, 51*(4), 364-380. https://doi.org/10.1111/radm.12471

Stepanyan, K., Littlejohn, A., & Margaryan, A. (2013). Sustainable e-learning: Toward a coherent body of knowledge. *Educational Technology & Society, 16*(2), 91–102.

Teng, C. I. (2010). Customization, immersion satisfaction, and online gamer loyalty. *Computers in Human Behavior, 26*(6), 1547–1554. https://doi.org/10.1016/j.chb.2010.05.029

Tesar, M. (2020). Towards a post-Covid-19 “new normality?”: Physical and social distancing, the move to online and higher education. *Policy Futures in Education, 18*(5), 556–559. https://doi.org/10.1177/1478210320935671

Treiblmaier, H., Madlberger, M., Knotzer, N., & Pollach, I. (2004). Evaluating personalization and customization from an ethical point of view: An Empirical study. *Proceedings of the Hawaii International Conference on System Sciences, 37*, 2831–2840.

Tumbas, S., Berente, N., Seidel, S., & Vom Brocke, J. (2015). The “Digital Façade” of rapidly growing entrepreneurial organizations. *2015 International Conference on Information Systems: Exploring the Information Frontier, ICIS 2015*.

Vlachopoulos, D. (2020). Covid-19: Threat or opportunity for online education? In *Higher Learning Research Communications, 10*(1). https://doi.org/10.18870/hlrc.v10i1.1179

Winter, S. G., & Szulanski, G. (2001). Replication as strategy. *Organization Science, 12*(6), 730–743. https://doi.org/10.1287/orsc.12.6.730.10084

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. https://doi.org/10.1016/j.compedu.2019.103599

Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). Sage Publications.

Zulfikar, M. W., bin Hashim, A. I., bin Ahmad Umri, H. U., & Dahlan, A. R. A (2018). A Business case for digital transformation of a Malaysian-based university. *2018 International Conference on Information and Communication Technology for the Muslim World (ICT4M)* (pp. 106–109). https://doi.org/10.1109/ICT4M.2018.00028

Hong, J.-C., Liu, X., Cao, W., Tai, K.-H., & Zhao, L. (2022). Effects of Self-Efficacy and Online Learning Mind States on Learning Ineffectiveness during the COVID-19 Lockdown. *Educational Technology & Society*, *25* (1), 142-154.

Effects of Self-Efficacy and Online Learning Mind States on Learning Ineffectiveness during the COVID-19 Lockdown

Jon-Chao Hong[1], Xiaohong Liu[2], Wei Cao[2], Kai-Hsin Tai[1] and Li Zhao[2*]
[1]Department of Industrial Education, National Taiwan Normal University, Taiwan // [2]School of Education Science, Nanjing Normal University, Nanjing, China // tcdahong@gmail.com // xiaohongliu1211@gmail.com // 2576901707@qq.com // star99xin@gmail.com // li.zhao@njnu.edu.cn
[*]Corresponding author

ABSTRACT: With the outbreak of COVID-19, more online learning has been adopted for distance learning. However, the effectiveness of online learning for those students engaged in it for the first time has not been discussed. This study aims to investigate perceived ineffectiveness of online learning and its antecedents related to cognitive and affective factors. Internet self-efficacy (ISE) and Self-efficacy of interacting with learning content (SEILC) were hypothesized to have a correlation with perceived ineffectiveness of online learning (PIOL) mediated by participants' Internet cognitive fatigue (ICF) and mind-unwandered, while ICF was hypothesized to have a correlation with mind-unwandered. Data of 251 students collected from high schools in China during the lockdown period of COVID-19 were subjected to confirmatory factor analysis via AMOS. Results indicated that participants' ISE and SEILC were positively related to mind-unwandered, but negatively related to ICF during online learning, while ICF was positively associated with PIOL. On the other hand, mind-unwandered was negatively associated with PIOL. Furthermore, students' ISE and SEILC indirectly affected their PIOL mediated by ICF. Findings suggest that an enhancement of learners' ISE and SEILC could have reduced the level of PIOL the first time that online learners experienced under the COVID-19 lockdown to promote their learning effectiveness. This understanding will be useful in case of another pandemic outbreak.

Keywords: E-learning, High school students, Internet cognitive fatigue, Mind-unwandered, Online learning

1. Introduction

More than 130 countries affected by the coronavirus outbreak have temporarily closed down offline educational facilities to contain the diffusion of COVID-19. To mitigate the immediate impact of school closures, the United Nations Educational, Scientific and Cultural Organization (UNESCO) has launched distance learning solutions (UNESCO, 2020). Most online courses are synchronous online lectures via Zoom or Tencent Meeting.

With COVID-19 disrupting our learning and lives without warning, students had some difficulties with this urgent online distance learning (Zainuddin et al., 2020). Particularly, many teachers or students had no previous experience of online teaching or learning due to the main form of education still being face-to face learning. Therefore, this large-scale online learning phenomenon was undoubtedly a new challenge for the vast majority of teachers and students. For example, some students were not able to use any internet-enabled devices to participate in their study at home or to connect to a mobile network. Differences in the speed of Internet access and the type of learning device may also cause fatigue for some learners (Carter Jr et al., 2020). Whether in face-to-face or online classes, the key to a student's academic success is engagement (Buelow et al., 2018). In the traumatic environment of the COVID-19 epidemic, many learners may not be in a suitable emotional state to focus on learning (Carter Jr et al., 2020). How to maximize academic achievement or learning outcomes of online learning has been the focus of educators and researchers (Yokoyama, 2019), with most studies comparing the overall effect of online learning with traditional learning, or exploring the correlation between learning outcomes and behavior in online learning (e.g., number of attendance and discussions) (Koç, 2017; Zheng et al., 2020). Considering that the ineffectiveness of online learning has not been extensively studied in the context of online learning during the COVID-19 lockdown, this study explores the correlates of learning ineffectiveness in the online learning context. In particular, the learning ineffectiveness mentioned in this study refers to the negative evaluation of learning effectiveness by students who experienced online learning during the COVID-19 lockdown (Hong et al., 2021).

The cognitive theory of multimedia learning (CTML) attempts to explain how multimedia instructional design can affect learners' cognitive processing and learning performance (Mayer, 2005). In the learning process, individuals' cognition is limited in the face of multimedia information, and they can only process a certain amount of information within a given time (Liu et al., 2018). CTML provides a foundation for understanding factors that both promote and inhibit the input attention of learners in online learning. For example, students' attention levels influence their learning effectiveness within a MOOC learning environment (Chang et al., 2019). Moreover, students' attention plays a mediating role between their self-efficacy and achievement (Sun & Yeh, 2017). Extending the CTML to online learning environments, this study evaluates studies on attention level related to Internet cognitive fatigue (Hong et al., 2015) (i.e., being considered as cognitive fatigue in online learning) and mind-unwandered (Siegel, 2016) during online learning predicted by self-efficacy and reflecting learning ineffectiveness.

Individual nursing students' perceptions of self-efficacy were found to play a key role in their adoption of behaviors and maintenance of better performance (Karabacak et al., 2019). However, adolescents seldom express positive values of others' actions, and are likely to be biased in their response tendencies (Soto et al., 2008; van Herk et al., 2004). For example, an acquiescent "worth to myself" response is a tendency to respond negatively to survey items which are related to others or systems (Daniel & Benish-Weisman, 2019). For example, participants usually face some difficulties that prevent them from feeling satisfied with participating in online courses (Rabin et al., 2020). Considering this, by adopting the opposite self-rating, learning ineffectiveness replaced learning effectiveness for high-school-student participants to self-evaluate their perceptions of their online learning performance. This study aimed to explore the correlates between those students' different types of self-efficacy, Internet cognitive fatigue and mind-unwandered during online learning, and to determine whether those factors had a strong association with the high school students' perceptions of online learning ineffectiveness during the COVID-19 lockdown.

2. Theoretical background

Drive theory can be used to explain various individual difference measures, including motivation, attitudes, and psychological interests (Bouchard Jr., 2016). According to drive theory, there are two noteworthy ways to involve individuals in activities to achieve certain goals: 1) competence and confidence, and 2) cognition and emotion (Hrtoňová et al., 2015). Considering this, this study included participants' self-efficacy and attentional factors related to reflecting meaningfully on the ineffectiveness of online learning.

2.1. Internet self-efficacy and interacting with learning content self-efficacy in the context of online learning

According to Bandura's (1977) concept of self-efficacy, which is an individual's belief in his/her ability to succeed, individuals will try to do what they believe they can do, will choose to perform activities according to their efficacy beliefs, and will put efforts into activities and persist when faced with obstacles based on estimates of their efficacy. Considering the concept of interaction between the environments, the structure, and the individuals, the most prominent framework of interaction in distance education includes learner-environment interaction and learner–content interaction (Moore, 2013). Self-efficacy can affect performance (Morfoot & Stanley, 2018). When relating self-efficacy to online learning, researchers have proposed various types of self-efficacy from different angles (Hodges, 2008). For example, Chu and Tsai (2009) highlighted a two-dimensional category and classified ISE into general Internet self-efficacy (GISE) and communication Internet self-efficacy (CISE). GISE showed the confidence in overcoming the fundamental challenges associated with the operation of the Internet, whereas CISE is related to the confidence in communicating and interacting with others through the Internet (Chu, 2010; Chu & Tsai, 2009). Considering this, learners' interacting ability and confidence in online learning, considered as two types of self-efficacy: Internet self-efficacy (ISE) (i.e., learner-online system interaction) and self-efficacy of interacting with learning content (SEILC) (i.e., learner-content interaction), were taken into account in this study.

Internet self-efficacy (ISE) refers to users' self-efficacy when interacting with a website, the system itself, and interactive content designed for users. ISE has been defined as an individual's belief in his/her ability to successfully use the Internet, and is considered as an important antecedent of the effects of e-learning (Eastin & LaRose, 2000; Jokisch et al., 2020). Additionally, with respect to interactive actions, content should have a strong relationship to information searching that has effects on learning self-efficacy (Jokisch et al., 2020). Regarding the interactive content in online learning systems, self-efficacy can achieve the confidence of information transfer between humans and computers (Hong et al., 2011). Accordingly, the two types of online learning self-efficacy: ISE and Self-efficacy of interacting with learning content (SEILC), were examined to understand how they affect participants' online learning, as mentioned above. Considering online learning during the COVID-19 lockdown in which students had to interact with transactional media and transactional content, this study explored how their ISE and SEILC interacted during their online learning was explored in this study.

2.2. Attention factors: Internet cognitive fatigue and mind-unwandered in the context of online learning

As a key factor in cognitive processing and human perception, attention can arouse individual's perception of events and lead to the cognitive processing needed for meaningful learning (Baars, 1997). Humans cannot respond to or process all of the environmental stimuli they encounter due to their limited attention capacities (Pashler, 1998). In online learning environments, learners obtain the information that meets their respective aims by performing online searches, which requires them to pay attention to the learning tasks (Wu & Xie, 2018). Therefore, within the online search paradigm, focused attention on tasks related to active cognitive control is related to current information processing.

Attention is usually treated as a dichotomy: inattention is considered as mind-wandering as a result of losing attention when performing a task, while the other option is full attention, which is postulated as mind-unwandered when attention is focused on a task. At different hierarchical levels of cognitive processing, mind-unwandered can keep attention focused on the external input and sustain cognitive processing (Schad et al., 2012). In addition, mind-unwandered refers to paying attention to one's thoughts and emotions, which make one's experiences on a moment-to-moment basis during the cognitive process (Siegel, 2016). Mind-unwandered therefore pertains to task-related thought which can probe on-task thoughts in online education contexts.

As for cognitive processing, attention failure can be assessed by comparing participants' response errors with the distractions that they experience when performing tasks, which results in cognitive fatigue (Geva et al., 2013). Cognitive fatigue is defined by DeLuca as "time-related deterioration in the ability to perform certain mental tasks" (2005, p. 38). Accordingly, Hong et al. (2015) extended this type of cognitive disability to the Internet world, as Internet cognitive fatigue (ICF), with the aim of exploring the effect of ICF on vocabulary memorization to support the argument that ICF represents a correlation or expression of a reduction in a learner's online performance. Cognitive fatigue undermines task performance, because participants will reduce their attention and its allocation to stimuli that are unrelated to the task (Was et al., 2019).

When attention shifts from a task due to distractions in the environment, or due to internal thoughts, attention failures will occur, leading to failures in intended actions (Unsworth et al., 2012). On the other hand, during activities that are demanding and which require concentration, if an individual has a high level of mind-unwandered, it often leads to better performance and accuracy (Hollis & Was, 2016). The idea here is that online learning is likely a common need in the COVID-19 lockdown for students who are engaging in an online course and are hence interacting with an online learning system and trying to stay on task. Accordingly, the inhibitor or promoter for maintaining attention in online courses will influence students' performance. This study therefore aimed to evaluate individual differences in Internet cognitive fatigue and mind-unwandered, while participants had lessons in an online learning system, and it analyzed how various factors influenced their learning performance.

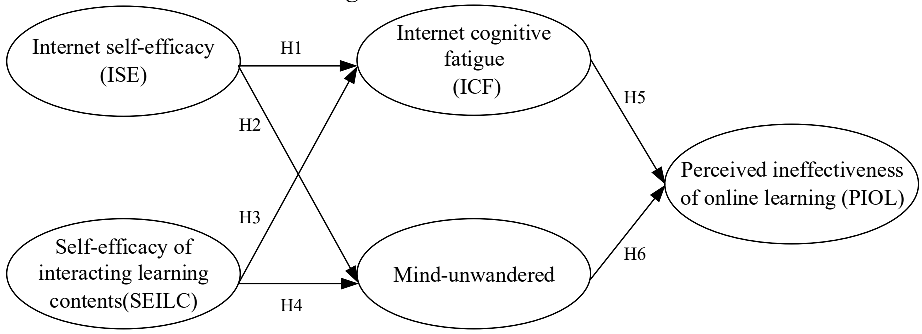
2.3. Research model and hypotheses

Gray (1982) conceptualized only two behavioral systems, the behavioral approach system (BAS) and the behavioral inhibition system (BIS). Oguchi and Takahashi (2019) suggested that the activated BIS predicts inattention and avoids the pursuit of desired goals; on the other hand, BAS drives more attention to the persistent pursuit of desired goals. Considering the learning potential effectiveness influenced by the behavioral system, BAS is activated by positive factor stimuli, while BIS is activated by negative factor stimuli; both stimuli affect learning performance (Chan & Tse, 2018). Accordingly, the present study focused on the attention level related to the deactivating factor: Internet cognitive fatigue as a BIS factor, and the activating factor, mind-unwandered, as a BAS factor, predicted by the positive psychological trait, self-efficacy, that reflects learning ineffectiveness in an online learning context. Thus, to explore the correlates between those factors, the present study referred to drive theory to develop a conceptual model, shown as follows (see Figure 1).

Working memory capacity is generally considered to be capable of processing information and of retaining it (van Merriënboer & Ayres, 2005). Cognitive ability affects the working memory capacity of the learners, who will then invest mental effort in maintaining attention to attain the learning that will enhance their performance outcomes (Kirschner et al., 2006). An individual with greater self-efficacy will experience a lower burden on working memory resources than an individual with less self-efficacy (Mayer et al., 2001). That is, self-efficacy enables learners to manage attention during practice in on-demand situations (Maertz Jr. et al., 2005). For example, Hong et al. (2016) posited that high levels of self-efficacy are related to low levels of Internet cognitive disability. In that sense, how ISE and SEILC affect Internet cognitive fatigue and mind-unwandered during online learning was hypothesized as follows:

H1: ISE is negatively related to students' ICF.
H2: ISE is positively related to students' mind-unwandered.
H3: SEILC is negatively related to students' ICF.
H4: SEILC is positively related to students' mind-unwandered.

Figure 1. Research model



Executive attention refers to the system that controls interference and resolves conflicts between possible reactions (Fan et al., 2002). According to Fougnie (2008), attention is mostly concerned with the manipulation of information during the learning process. Several studies have confirmed that the control of attention is strongly related to performance scope (Shipstead et al., 2016). For example, Musso et al. (2019) partially demonstrated the differential levels of cognitive processes that affect the prediction of mathematics performance. In addition, self-evaluation affects academic performance, both directly and indirectly mediated by cognitive ability (Demetriou et al., 2020), providing a foundation to further explore cognitive performance. Accordingly, the interaction effects between different types of attention: Internet cognitive fatigue and mind-unwandered on performance tasks, were hypothesized as follows:

H5: ICF is positively related to students' PIOL.
H6: Mind-unwandered is negatively related to students' PIOL.

In an online learning environment, students' self-efficacy is critical to improve learning performance. Students in online learning environments have a higher dropout rate (Bawa, 2016), and this dropout rate is related to students' low self-efficacy (Lee & Choi, 2011). In online learning, because students are required to mentally combine redundant information or integrate different sources of information, unnecessary working memory will be increased (Schmeck et al., 2015; Sweller et al., 2019). For example, in the context of online learning, if students waste too much time searching for information, then de-motivation tends to occur (Simunich et al., 2015). Therefore, the present study considered that the two types of self-efficacy (ISE and SEILC) would indirectly affect students' PIOL by affecting their Internet cognitive fatigue and mind-unwandered,. The following hypotheses were thus proposed:

H7a: The two types of self-efficacy are negatively related to PIOL mediated by ICF.
H7b: The two types of self-efficacy are negatively related to PIOL mediated by mind-unwandered.

3. Method

3.1. Data collection and participants

In this study, random sampling was adopted and data were collected using online questionnaires which were administered during the COVID-19 lockdown period of April 20-30, 2020. The data were collected through a web-based survey of 279 students from high schools in Jiangsu province, China. Participants took part in the online survey voluntarily and anonymously. However, 28 questionnaires were removed due to missing values or because the response time was too short. The remaining 251 data from the questionnaire were analyzed.

The participants consisted of 95 boys (37.8%) and 156 girls (62.2%). The students were aged between 15 and 18 years (M = 16.87, SD = .95); 39 (15.5%) reported that they spent less than 2h/day on online courses, 125 (49.8%) reported that they spent 2h-4h, 61 (24.3%) spent 4h-6h, while the remaining 26 (10.4%) reported spending more than 6h. As for the number of online courses the participants had attended in the current semester, 30 (12.0%) had attended less than 3, 203 (80.9%) had attended 4-6, 14 (5.5%) had attended 7-9, and the remaining four

(1.6%) participants had attended 10-12 online courses. Most of the participants (Frequency = 67, 26.7%) took courses online between 71% and 80% of the time in the current semester.

3.2. Instruments

The items of five constructs were adapted from previous studies and were created by having the original items professionally translated into Chinese. Face validity was conducted by research experts. Finally, a 5-point Likert scale was employed (i.e., ranging from 1 indicating *strongly disagree* to 5 indicating *strongly agree*), and the reliability of the constructs was subsequently tested. After omitting the items with low factor loadings or which were highly correlated with other items in the research model, final constructs showed good composite reliability, internal consistency reliability, and convergent validity (as shown in Table 1).

3.2.1. Internet self-efficacy measurement

This Internet self-efficacy scale was originally developed by Eastin and LaRose (2000), to assess the undergraduate students' Internet self-efficacy. In the context of ISE, consistency in appearance, control, and function of the website is important to the user (Cheng & Tsai, 2011). Accordingly, six items were designed in this study; exemplary items include: "I am confident in successfully dealing with the emergent problems of human-computer interaction in online learning" and "If I come across any trouble while using a website to learn, I have confidence in overcoming it."

3.2.2. Self-efficacy of interacting with learning content measurement

This study integrated Kuo's (2010) and Kao and Tsai's (2009) scales to develop the Self-Efficacy of Interacting with Learning Content measurement. All items were reviewed by two experts in online learning. Thus, six items were designed for this study; two example items are: "I have the confidence to understand new content on an e-learning platform" and "If I come across difficult content in e-learning, I have confidence in my ability to learn it well."

3.2.3. Internet cognitive fatigue measurement

This scale was adapted from Hong et al. (2015). It was originally developed to measure cognitive fatigue from time-on-task in terms of concentration, attention, memory, perception and motor control, and to evaluate the task-specific mistakes as time-related degradation in ability. Thus, five items were designed in this study; exemplary items include: "I lose concentration very quickly during online learning" and "I reach attention deficit very quickly during online learning."

3.2.4. Mind-unwandered measurement

The mind-unwandered scale was originally developed by Brown and Ryan (2003), to measure participants' general tendency to pay attention to assessing natural propensity and to focus on the current moment. Accordingly, the state of mind-unwandered as being fully attentive to present internal and external stimuli was considered when designing the questionnaire items. Thus, eight items were designed for this study, all of which were reviewed by two experts in online learning. Exemplary items include: "When studying online, I can follow the teacher's teaching steps even if I am away from the teacher" and "When I'm learning online, I don't leave the learning interface to do things that aren't related to what I'm learning in class."

3.2.5. Perceived ineffectiveness of online learning measurement

The Perceived Ineffectiveness of Online Learning scale was originally developed by Hong et al. (2021) to measure college students' perceived learning ineffectiveness. Six items were designed for this study; exemplary items are: "Since learning online, my learning efficiency has decreased" and "Since learning online, the quality of my homework has gotten worse."

3.3. Reliability and validity analysis

First, items with factor loading values less than 0.5 in each construct were deleted in each construct. After conducting CFA, items with the highest residual value in each construct were deleted until those CFA values reached the threshold suggested by Hair et al. (2019). The measurement model exhibited a good fit, with χ^2 = 132.276, df = 109, p < .001, χ^2/df = 1.214, GFI = .944, NFI = .952, CFI = .991, and RMSEA = .029. Hence, 22 remaining items were kept for further analysis, including three items each for ISE, SEILC, and ICF, and four each for mind-unwandered and PIOL.

Second, the internal and composite reliabilities of the questionnaire were analyzed. George and Mallery (2003) stated that if the Cronbach's alpha coefficient is greater than 0.7, it means that internal consistency is high, and reliability is high. The composite reliability (CR) over 0.70 indicates good external reliability (Hair et al., 2019). Table 1 displays that CR and Cronbach's alpha were both above 0.7, with CR ranging from .765 to .943 and Cronbach's alpha ranging from .764 to .942, indicating that the Cronbach's alpha and CR values of all the constructs met the threshold.

Third, convergent validity is determined by the factor load (FL) and average variable extraction (AVE) of each observed variable. The FL and AVE for each observed variable should be higher than 0.5 based on George and Mallery (2003) and Hair et al. (2019). Table 1 shows that the AVE of each construct was more than .50 (ranging from .522 to .805), and the FL of each construct was greater than .50 (ranging from .721 to .896). In sum, the convergent validity of each construct was acceptable.

Table 1. Reliability and validity analysis

Variables	*M*	*SD*	Cronbach's *α*	CR	AVE	FL
Threshold	--	--	> 0.7	> 0.7	> 0.5	> 0.5
Internet self-efficacy	2.426	0.783	0.786	0.789	0.557	0.743
Self-efficacy of interacting with the learning content	3.988	0.720	0.729	0.800	0.573	0.755
Internet cognitive fatigue	3.714	0.623	0.764	0.765	0.522	0.721
Mind-unwandered	3.836	0.637	0.898	0.901	0.695	0.831
Perceived ineffectiveness of online learning	2.675	1.054	0.942	0.943	0.805	0.896

3.4. Data analysis

Descriptive statistics of participants' information and the reliability and validity of the questionnaire were obtained in the current study by using SPSS (version 20.0). Moreover, we used first-order confirmatory factor analysis (CFA) to confirm the item suitability of the measuring questionnaire. Afterward, model-fit indexes of the measurement items were used to verify the measurement model. Structural equation modeling (SEM) was then conducted to assess the hypothetical structural model via AMOS (version 22.0).

4. Results

4.1. Model fit analysis

The model fit and statistical significance of the hypothesized path among the five potential variables were examined to test the structural model. The standardized regression weight, item communalities, and model-fit indexes of the measurement items were applied to identify the structural validity of the measurement model. Various measures were conducted to assess the fit of the models, such as the root mean square error of approximation (RMSEA), the goodness of fit index (GFI), the normed fit index (NFI), the comparative fit index (CFI), and chi-square normalized by degree of freedom (Chi-square/df).

The χ^2 of this study is 274.378 and the degree of freedom (df) is 113, which makes χ^2/df equal to 2.428 The resulting ratio is less than 3, which is regarded as being indicative of a good fit (Kline, 2010). The RMSEA value below .08 is considered to be a good fit. On the other hand, a GFI value below .08 means a good model fit (Hair et al., 2019). Moreover, Kline (2010) suggested that the AGFI value has to surpass the threshold value of .80. In the present study, RMSEA was .076, GFI was .901, and AGFI was .875, all meeting the threshold values. Additionally, the Normed Fit Index (NFI) was .901, the Non-Normed Fit Index (NNFI) was .926, the Comparative Fit Index (CFI) was .939, and the Incremental Fit Index (IFI) was .939; therefore, the present model fits were all above .90, indicating a good fit (Kline, 2010). Moreover, PNFI and Parsimonious Goodness of Fit

Index (PGFI) were .749 and .665, which passed the suggested threshold value of .5 (Hair et al., 2019). These indicators show that the hypothesis model proposed in this study has good fitness.

4.2. Path analysis

To test the six hypotheses, AMOS was used to calculate the correlation coefficient among the five latent constructs and the research model's explanatory power. The standardized path coefficients of the hypothesized model are shown in Table 2 and Figure 2. The results indicate that Hypotheses 1, 2, 3, 4, 5 and 6 were all supported. The ISE was positively related to SEILC and mind-unwandered (β = 0.567, t = 7.133***; β = 0.326, t = 4.980**). The ISE and SEILC were negatively related to ICF (β = -0.374, t = -4.402***; β = -0.423, t = -4.954***). Moreover, ICF was positively related to PIOL (β = 0.350, t = 4.287***), and mind-unwandered was negatively associated with PIOL (β = -0.211, t = -3.041**).

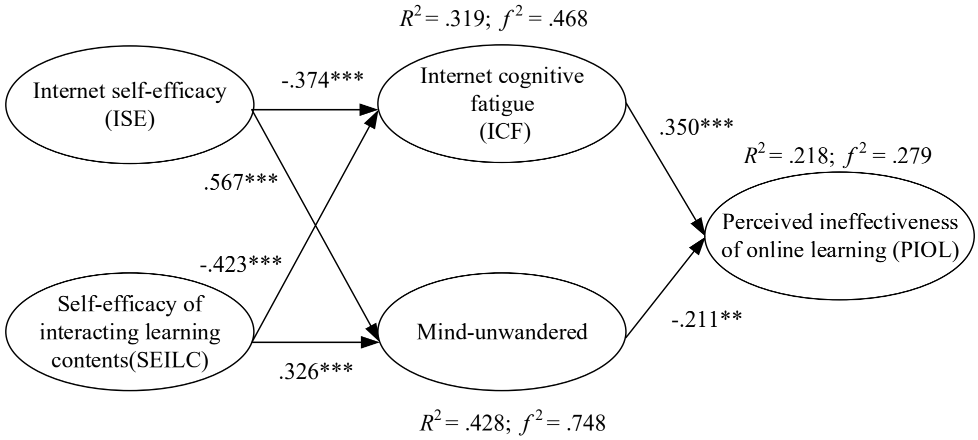
The coefficient of determination (R^2) represents the overall impact of the exogenous variable on the endogenous variable. R^2 values higher than 0.6 are considered to have a high impact effect, 0.3-0.6 are considered medium, and less than 0.3 is considered as having a low impact effect (Sanchez, 2013). Those R^2 values in Figure 2 indicate that ISE and SEILC had a medium impact on ICF and mind-unwandered, and the effect of ICF and mind-unwandered on PIOL was low. In addition, effect size (Cohen's f^2) was proposed by Cohen (1988), where f^2 values greater than 0.8, between 0.2 and 0.8, and less than 0.2 can be considered as large, medium and small, respectively. As shown in Figure 2, the explanatory power of ISE and SEILC on ICF was 31.9% (f^2 =.468), and on mind-unwandered it was 42.8% (f^2 =.748). The explanatory variance of ICF and mind-unwandered on PIOL was 21.8% (f^2 =.279). Hence, the six variables in this study have good predictive power (Hair et al., 2019).

Table 2. Coefficients of the hypothesized model

Hypothesis	Path	Standardized coefficient (β)	*S.E.*	*t*	Supported?
H1	ISE→ ICF	-0.374	0.103	-4.402***	Yes
H2	ISE→ Mind-unwandered	0.567	0.116	7.133***	Yes
H3	SEILC→ ICF	-0.423	0.099	-4.954***	Yes
H4	SEILC→ Mind-unwandered	0.326	0.091	4.98***	Yes
H5	ICF→PIOL	0.35	0.137	4.287***	Yes
H6	M→PIOL	-0.211	0.097	-3.041**	Yes

Note. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

Figure 2. Verification of the research model



4.3. Indirect effects of SEILC and ISE on PIOL mediated by two types of attention

To provide additional evidence to explore whether the indirect effects contained in the research model are significant, 1,000 resample bootstrappings were performed in this study. The bootstrapping results are shown in Table 3, which provides the un-standardized coefficient and upper and lower bound of 95% confidence intervals. It can be observed that the bootstrapping confidence intervals of indirect effects did not comprise zero in the two paths, including ISE→ICF→PIOL (95% CI= [-.398, -0.044]) and SEILC→ICF→PIOL (95%CI = [-0.408, -0.046]). Therefore, ISE and SEILC were negatively related to PIOL mediated by ICF, revealing that H7a was supported. Mind-unwandered did not mediate the effect from ISE and SEILC to PIOL, due to the bootstrapping confidence interval of indirect effects which contained zero, indicating that H7b was unsupported.

Table 3. Bootstrapping results

Model paths	Un-standardized coefficient	95% CI	
		Lower bound	Upper bound
Indirect effect			
ISE→ ICF→ PIOL	-.268	-0.398	-0.044
ISE→ Mind-unwandered→ PIOL	-.244	-0.359	0.077
SEILC→ ICF→ PIOL	-.288	-0.408	-0.046
SEILC→ Mind-unwandered→ PIOL	-.133	-0.205	0.038

5. Discussion

Considering the potential learning effectiveness influenced by the behavioral system, BAS was activated by positive factor stimuli, and BIS was activated by negative factor stimuli, with both stimuli affecting learning performance (Chan & Tse, 2018). Accordingly, the present study is focused on the attention level related to the BIS factor: Internet cognitive fatigue, and the BAS factor: mind-unwandered, predicted by the positive psychological trait: self-efficacy, that reflects the perception of learning ineffectiveness in an online learning context. Basically, this behavioral system provided a multidimensional model for understanding online learning with an emphasis on student focus factors (i.e., Mind-unwandered and ICF) and self-efficacy (i.e., ISE and SEILC). The results of this study help us to understand the students' perceived ineffectiveness of online learning during the COVID-19 lockdown.

Working memory capacity is generally considered to be capable of processing information and of retaining it (van Merriënboer & Ayres, 2005). With high working memory capacity, an individual will have greater self-efficacy; on the other hand, an individual with less self-efficacy will experience a lower burden on working memory resources (Mayer et al., 2001). Self-efficacy of interacting with learning content is another predictor of students' participation in online learning, due to its being able to build trust in the interaction between the user and the computer (Hong et al., 2011). In addition, ISE is an important predictor of students' participation in the online learning environment (Kuo et al., 2014). The present study further confirms this point, and the results suggest that ISE and SEILC show positive effects on students' mind-unwandered and negative effects on their ICF. H1 and H3 were hence negatively supported, and H2 and H4 were positively supported.

As cognitive ability affects the working memory capacity of the learners, they will then invest mental efforts in paying attention to attaining the learning that will enhance their performance outcomes (Kirschner et al., 2006). For example, cognitive fatigue is usually accompanied by loss of concentration (Was et al., 2019), and this relationship still exists in Internet cognitive fatigue (Hong et al., 2015). The results of the present study verified that ICF can positively predict perceived learning ineffectiveness, revealing that H5 was positively supported.

In an online learning environment, mind-unwandered is an important prerequisite for students to participate in learning activities. However, mind-unwandered is easily interfered with by environmental and personal concerns, especially when learners need to focus on multitasking (Miller et al., 2020; Sana et al., 2013). The results of the current study showed that the negative effect of mind-unwandered on the students' perceived ineffectiveness of online learning was significant. This finding supports Was's et al. (2019) view that mind-wandering is detrimental to learners' learning of course content, and potentially damaging to their learning performance in an online learning environment, showing that H6 was negatively supported.

Perceived ineffectiveness of online learning, as the cognition of students in online learning, is also a factor that should be captured as part of learners' learning outcomes (Ruhland & Brewer, 2001). In summary, ISE and SEILC have an indirect relationship with students' perceived ineffectiveness of online learning, mediated by Internet cognitive fatigue. Therefore, this result was supported by several researchers' views that there is a correlation between self-efficacy and students' learning performance (e.g., Huang & Mayer, 2018; Pellas, 2014). Thus, H7a was supported. In addition, the two types of self-efficacy are positively correlated with attention, and attention is negatively correlated with learning effect. However, mind-unwandered does not play a mediator role in the indirect effect from the two types of self-efficacy to PIOL (and so H7b was not supported).

6. Conclusions

How to promote the effectiveness of online learning is important in the period of the pandemic lockdown. To understand how high school students perceive their efficacy of interacting with an online learning system and content, and their attentional states when interacting with online learning, which is then reflected in their perception of learning ineffectiveness, this study distinguished two types of self-efficacy (Internet self-efficacy and self-efficacy of interacting with learning content) in the context of students' online learning, while exploring how these two types of self-efficacy affect the students' perceived ineffectiveness of online learning as mediated by Internet cognitive fatigue and mind-unwandered. The results provided evidence to show that high school students' perceived ineffectiveness of online learning can be reduced when their mind-unwandered is improved upon and when their cognitive fatigue is reduced. In addition, students' PIOL was indirectly affected by their ISE and SEILC, mediated by ICF.

6.1. Implications

The theoretical contribution of this study is to prove that Internet self-efficacy and self-efficacy in interacting with learning content do extend to the online learning environment, and it was validated that these two kinds of self-efficacy will indirectly influence the students' perceived ineffectiveness of online learning.

The practical contribution of this research is that the findings can provide some guidance to instructors in order to improve their online learning classes. For example, teachers should provide guidance for those students with low Internet self-efficacy and self-efficacy of interacting with learning content. They can provide reminders when students' minds start wandering, and increase learners' interaction within the teaching tool to prolong their mind-unwandered (Ha & Im, 2020; Sun & Yeh, 2017). In addition, teachers can also design their own methods to strengthen the interactivity and collaboration of online learning activities (Liu et al., 2021), helping to reduce isolation and lack of interaction between students in distance online learning, thus improving students' attention.

Finally, it would also be beneficial for teachers to improve the students' ISE and SEILC in order to save online learning time. Enhancing students' mind-unwandered and reducing their ICF will in turn increase their online learning effectiveness.

6.2. Limitations and future study

Although the present study provides some important contributions to the literature, there are several limitations which should be recognized. First, the causal relationship among the observed variables cannot be determined because of the cross-sectional survey. Second, the data of this study were collected in one province of China by random sampling, which did not cover high schools of different levels and therefore cannot represent all Chinese high school students. More and larger representative samples will be needed in the future to assess the extent to which the findings are applicable to other population groups and other countries to confirm the hypotheses of the present study.

Another limitation is that the participants had to receive online learning to avoid the spread of the COVID-19 outbreak. It is unclear whether students' features would produce the same findings in different settings or stages.

In addition, other factors not covered in this study may also affect students' perceived ineffectiveness of online learning, such as self-regulated learning, learning motivation, learning satisfaction, online interaction quality, and academic procrastination. Future studies might consider adding other factors to future studies that may have effects on perceived ineffectiveness of online learning.

The present study proposed a research model to explore the indirect effect between ISE/SEILC, and PIOL mediated by ICF and mind-unwandered, and there were negative predictions. However, we did not test the direct effect between ISE/SEILC and PIOL; future studies may focus on examining their correlation.

Acknowledgement

This research was sponsored by the National Social Science Foundation (No. BCA200093).

References

Baars, B. J. (1997). Some essential differences between consciousness and attention, perception, and working memory. *Consciousness and Cognition, 6*(2–3), 363–371. https://doi.org/10.1006/ccog.1997.0307

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

Bawa, P. (2016). Retention in online courses: Exploring issues and solutions—A Literature review. *Sage Open, 6*(1),1–11. https://doi.org/10.1177/215824401562177

Bouchard Jr., T. J. (2016). Experience producing drive theory: Personality "writ large." *Personality and Individual Differences, 90,* 302-314. https://doi.org/10.1016/j.paid.2015.11.007

Brown, K. W., & Ryan, R. M. (2003). The Benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*(4), 822–848. https://doi.org/10.1037/0022-3514.84.4.822

Buelow, J. R., Barry, T., & Rich, L. E. (2018). Supporting learning engagement with online students. *Online Learning, 22*(4), 313–340. https://doi.org/10.24059/olj.v22i4.1384

Carter Jr, R. A., Rice, M., Yang, S. & Jackson, H. A. (2020). Self-regulated learning in online learning environments: Strategies for remote learning Strategies. *Information and Learning Sciences, 121*(5/6), 321-329. https://doi.org/10.1108/ILS-04-2020-0114

Chan, S. H. W., & Tse, S. (2018). An Explorative study on coping flexibility with behavioral approach system activating stimuli: A Comparison of people with and without bipolar disorder. *Psychiatry Research, 269*, 399–407. https://doi.org/10.1016/j.psychres.2018.08.084

Chang, J. J., Lin, W. S., & Chen, H. R. (2019). How attention level and cognitive style affect learning in a MOOC environment? Based on the perspective of brainwave analysis. *Computers in Human Behavior, 100*, 209–217. https://doi.org/10.1016/j.chb.2018.08.016

Cheng, K. H., & Tsai, C. C. (2011). An Investigation of Taiwan University students' perceptions of online academic help seeking, and their online learning self-efficacy. *The Internet and Higher Education, 14*(3), 150–157. https://doi.org/10.1016/j.iheduc.2011.04.002

Chu, J. C. (2010). How family support and internet self-efficacy influence the effects of e-learning among higher aged adults – Analyses of gender and age differences. *Computers & Education, 55*(1), 255-264. https://doi.org/10.1016/j.compedu.2010.01.011

Chu, R. J.-C., & Tsai, C. C. (2009). Self-directed learning readiness, Internet self-efficacy and preferences towards constructivist Internet-based learning environments among higher-aged adults. *Journal of Computer Assisted Learning, 25*(5), 489–501. https://doi.org/10.1111/j.1365-2729.2009.00324.x.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Daniel, E., & Benish-Weisman, M. (2019). Value development during adolescence: Dimensions of change and stability. *Journal of Personality, 87*(3), 620–632. https://doi.org/10.1111/jopy.12420

DeLuca, J. (2005). Fatigue, cognition, and mental effort. In J. Deluca (Eds.), *Fatigue as a window to the human brain* (pp. 37–57). The MIT Press.

Demetriou, A., Kazi, S., Makris, N., & Spanoudis, G. (2020). Cognitive ability, cognitive self-awareness, and school performance: From childhood to adolescence. *Intelligence, 79*, 101432. https://doi.org/10.1016/j.intell.2020.101432

Eastin, M. S., & LaRose, R. (2000). Internet self-efficacy and the psychology of the digital divide. *Journal of Computer-Mediated Communication, 6*(1), JCMC611. https://doi.org/10.1111/j.1083-6101.2000.tb00110.x

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14*(3), 340–347. https://doi.org/10.1162/089892902317361886

Fougnie, D. (2008). The Relationship between attention and working memory. In N. B. Johansen (Eds.), *New research on short-term memory* (pp. 1–45). Nova Science Publishers.

George, D. & Mallery P. (2003). *SPSS for Windows step by step: A Simple guide and reference.* (11.0 update, 4th ed.). Allyn and Bacon.

Geva, R., Zivan, M., Warsha, A., & Olchik, D. (2013). Alerting, orienting or executive attention networks: Differential patters of pupil dilations. *Frontiers in Behavioral Neuroscience, 7*, 145. https://doi.org/10.3389/fnbeh.2013.00145

Gray, J. A. (1982). *The Neuropsychology of anxiety: An Enquiry into the functions of the septo-hippocampal system*. Oxford University Press.

Ha, Y. & Im, H. (2020). The Role of an interactive visual learning tool and its personalizability in online learning: Flow experience. *Online Learning, 24*(1), 205-226. https://doi.org/10.24059/olj.v24i1.1620

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage.

Hodges, C. B. (2008). Self-efficacy in the context of online learning environments: A Review of the literature and directions for research. *Performance Improvement Quarterly, 20*(3-4), 7-25. https://doi.org/10.1002/piq.20001

Hollis, R. B., & Was, C. A. (2016). Mind wandering, control failures, and social media distractions in online learning. *Learning and Instruction, 42*, 104-112. https://doi.org/10.1016/j.learninstruc.2016.01.007

Hong, J. C., Hwang, M. Y., Chang, S. W., Tai, K. H., Kuo, Y. C., & Tsai, Y. H. (2015). Internet cognitive failure and fatigue relevant to learners' self-regulation and learning progress in English vocabulary with a calibration scheme. *Journal of Computer-assisted Learning, 31*(5), 450-461. https://doi.org/10.1111/jcal.12103

Hong, J. C., Hwang, M. Y., Szeto, S. Y. E., Tsai, C. R., Kuo, Y. C., & Hsu, W. Y. (2016). Internet cognitive failure relevant to self-efficacy, learning interest, and satisfaction with social media learning. *Computers in Human Behavior, 55*, 214-222. https://doi.org/10.1016/j.chb.2015.09.010

Hong, J. C., Hwang, M. Y., Hsu, H. F., Wong, W. T., & Chen, M. Y, (2011). Applying the technology acceptance model in a study of the factors affecting usage of the Taiwan digital archives system. *Computers & Education, 57*(3), 2086-2094. https://doi.org/10.1016/j.compedu.2011.04.011

Hong, J. C., Lee, Y. F., & Ye, J. H. (2021). Procrastination predicts online self-regulated learning and online learning ineffectiveness during the coronavirus lockdown. *Personality and Individual Differences, 174*(7), 110673. https://doi.org/10.1016/j.paid.2021.110673

Hrtoňová, N., Kohout, J., Rohlíková, L., & Zounek, J. (2015). Factors influencing acceptance of e-learning by teachers in the Czech Republic. *Computers in Human Behavior, 51,* 873-879. https://doi.org/10.1016/j.chb.2014.11.018

Huang, X., & Mayer, R. E. (2018). Adding self-efficacy features to an online statistics lesson. *Journal of Educational Computing Research, 57*(4), 1003-1037. https://doi.org/10.1177/0735633118771085

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, 111*, 106408. https://doi.org/10.1016/j.chb.2020.106408

Kao, C.-P., & Tsai, C.-C. (2009). Teachers' attitudes toward web-based professional development, with relation to Internet self-efficacy and beliefs about web-based learning. *Computers & Education, 53*(1), 66-73. https://doi.org/10.1016/j.compedu.2

Karabacak, U., Unver, V., Ugur, E., Kocatepe, V., Ocaktan, N., Ates, E., & Uslu, Y. (2019). Examining the effect of simulation based learning on self-efficacy and performance of first-year nursing students. *Nurse Education in Practice, 36*, 139–143. https://doi.org/10.1016/j.nepr.2019.03.012

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

Kirschner, P., Sweller, J., & Clark, R. E. (2006). Why minimally guided learning does not work: An Analysis of the failure of discovery learning, problem-based learning, experiential learning and inquiry-based learning. *Educational Psychologist, 41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Koç, M. (2017). Learning analytics of student participation and achievement in online distance education: A Structural equation modeling. *Educational Sciences: Theory & Practice, 17*(6), 1893–1910. https://doi.org/10.12738/estp.2017.6.0059

Kuo, Y.-C. (2010). *Interaction, internet self-efficacy, and self-regulated learning as predictors of student satisfaction in distance education courses* (Doctoral dissertation). Utah State University. https://doi.org/10.26076/157a-7cf6

Kuo, Y. C., Walker, A., Schroder, K. E. E., & Belland, B. R. (2014). Interaction, internet self-efficacy, and self-regulated learning as predictors of student satisfaction in online education courses. *The Internet and Higher Education, 20*, 35–50. https://doi.org/10.1016/j.iheduc.2013.10.001

Lee, Y. & Choi, J. (2011). A Review of online course dropout research: Implications for practice and future research. *Educational Technology Research and Development, 59*, 593–618. https://doi.org/10.1007/s11423-010-9177-y

Liu, H., Shao M., Liu X., & Zhao L. (2021). Exploring the influential factors on readers' continuance intentions of e-book APPs: Personalization, usefulness, playfulness, and satisfaction. *Frontiers in Psychology*, *12*, 640110. https://doi.org/10.3389/fpsyg.2021.640110

Liu, Y., Jang, B. G., & Roy-Campbell, Z. (2018). Optimum input mode in the modality and redundancy principles for university ESL students' multimedia learning. *Computers & Education, 127*, 190–200. https://doi.org/10.1016/j.compedu.2018.08.025

Maertz Jr., C. P., Bauer, T. N., Mosley Jr., D. C., Posthuma, R. A., & Campion, M. A. (2005). Predictors of self-efficacy for cognitive ability employment testing. *Journal of Business Research, 58*(2), 160–167. https://doi.org/10.1016/S0148-2963(03)00111-5

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Eds.), *The Cambridge handbook of multimedia learning* (pp. 31-48). Cambridge University Press.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187–198. https://doi.org/10.1037/0022-0663.93.1.187

Miller, M. D., Doherty, J. J., Butler, N. M., & Coull, W. G. (2020). Changing counterproductive beliefs about attention, memory, and multitasking: Impacts of a brief, fully online module. *Applied Cognitive Psychology, 34*(3), 710-723. https://doi.org/10.1002/acp.3662

Moore, M. G. (2013). The Theory of transactional distance. In M. G. Moore (Ed.), *Handbook of Distance Education* (3rd ed., pp. 66–85). Routledge.

Morfoot, C., & Stanley, H. (2018). Simulation-based education for neonatal skills training and its impact on self-efficacy in post-registration nurses. *Infant, 14*(2), 77–81. https://www.infantjournal.co.uk/pdf/inf_080_ain.pdf

Musso, M. F., Boekaerts, M., Segers, M., & Cascallar, E. C. (2019). Individual differences in basic cognitive processes and self-regulated learning: Their interaction effects on math performance. *Learning and Individual Differences, 71*, 58–70. https://doi.org/10.1016/j.lindif.2019.03.003

Oguchi, M., & Takahashi, F. (2019). Behavioral inhibition/approach systems constitute risk/protective pathways from ADHD symptoms to depression and anxiety in undergraduate students. *Personality and Individual Differences, 144*, 31–35. https://doi.org/10.1016/j.paid.2019.02.033

Pashler, H. E. (1998). *The Psychology of attention*. The MIT Press.

Pellas, N. (2014). The Influence of computer self-efficacy, metacognitive self-regulation and self-esteem on student engagement in online learning programs: Evidence from the virtual world of Second Life. *Computers in Human Behavior, 35*, 157–170. https://doi.org/10.1016/j.chb.2014.02.048

Rabin, E., Henderikx, M., Kalman, Y. M., & Kalz, M. (2020). What are the barriers to learners' satisfaction in MOOCs and what predicts them? The role of age, intention, self-regulation, self-efficacy and motivation. *Australasian Journal of Educational Technology, 36*(3), 119–131. https://doi.org/10.14742/ajet.5919

Ruhland, S. K., & Brewer, J. A. (2001). Implementing an assessment plan to document student learning in a two-year technical college. *Journal of Vocational Education Research, 26*(2), 141-171. https://doi.org/10.5328/JVER26.2.141

Sana, F., Weston, T., & Cepeda, N. J. (2013). Laptop multitasking hinders classroom learning for both users and nearby peers. *Computers & Education, 62*, 24–31. https://doi.org/10.1016/j.compedu.2012.10.003

Sanchez, G. (2013). *PLS path modeling with R*. (Trowchez Editions). https://www.gastonsanchez.com/PLS_Path_Modeling_with_R.pdf

Schad, D. J., Nuthmann, A., & Engbert, R. (2012). Your mind wanders weakly, your mind wanders deeply: Objective measures reveal mindless reading at different levels. *Cognition, 125*(5), 179–194. https://doi.org/10.1016/j.cognition.2012.07.004

Schmeck, A., Opfermann, M., van Gog, T., Paas, F., & Leutner, D. (2015). Measuring cognitive load with subjective rating scales during problem solving: Differences between immediate and delayed ratings. *Instructional Science, 43*(1), 93–114. https://doi.org/10.1007/s11251-014-9328-3

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science, 11*(6), 771–799. https://doi.org/10.1177/1745691616650647.

Siegel, R. D. (2016). *Positive psychology: Harnessing the power of happiness, mind-unwandered, and inner strength*. Boston, MA: Harvard Health Publications. http://www.sobtell.com/images/questions/1506616731-positive-psychology-harnessing-the-power-of-happiness-mindfulness-and-inner-strength-harvard-health.pdf

Simunich, B., Robins, D. B., & Kelly, V. (2015). The Impact of findability on student motivation, self-efficacy, and perceptions of online course quality. *American Journal of Distance Education, 29*(3), 174–185.https://doi.org/10.1080/08923647.2015.1058604

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The Developmental psychometrics of big five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology, 94*(4), 718–737. https://doi.org/10.1037/0022-3514.94.4.718

Sun, J. C.Y., & Yeh, K. P. C. (2017). The Effects of attention monitoring with EEG biofeedback on university students' attention and self-efficacy: The Case of anti-phishing instructional materials. *Computers & Education, 106*, 73-82. https://doi.org/10.1016/j.compedu.2016.12.003

Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

UNESCO. (2020). *Distance learning solutions*. https://en.unesco.org/covid19/educationresponse/solutions

Unsworth, N., McMillan, B. D., Brewer, G. A., & Spillers, G. J. (2012). Everyday attention failures: An Individual differences investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(6), 1765–1772. https://doi.org/10.1037/a0028075

van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology, 35*(3), 346–360. https://doi.org/10.1177/0022022104264126

van Merriënboer, J. J., & Ayres, P. (2005). Research on cognitive load theory and its design implications for e-learning. *Educational Technology Research & Development, 53*(3), 5–13. https://doi.org/10.1007/BF02504793

Was, C. A., Hollis, R. B., & Dunlosky, J. (2019). Do students understand the detrimental effects of mind wandering during online learning. *Computers & Education, 135*, 113–122. https://doi.org/10.1016/j.compedu.2019.02.020

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. https://doi.org/10.1016/j.chb.2018.07.008

Yokoyama, S. (2019) Academic self-efficacy and academic performance in online learning: A Mini review. *Frontiers in Psychology, 9*, 2794. https://doi.org/10.3389/fpsyg.2018.02794

Zainuddin, Z., Perera, C. J., Haruna, H., & Habiburrahim, H. (2020). Literacy in the new norm: Stay-home game plan for parents. *Information and Learning Sciences, 121*(7/8), 645-653. https://doi.org/10.1108/ILS-04-2020-0069

Zheng, B., Lin, C. H., & Kwon, J. B. (2020). The Impact of learner-, instructor-, and course-level factors on online learning. *Computers & Education*, *150*(1), 103851. https://doi.org/10.1016/j.compedu.2020.103851

Lin, C-H., Wu, W.-H., & Lee, T.-N. (2022). Using an Online Learning Platform to Show Students' Achievements and Attention in the Video Lecture and Online Practice Learning Environments. *Educational Technology & Society*, *25* (1), 155-165.

Using an Online Learning Platform to Show Students' Achievements and Attention in the Video Lecture and Online Practice Learning Environments

Chih-Hung Lin[1], Wun-Hau Wu[2] and Tsu-Nan Lee[3*]

[1]Master Program in Mathematics and Science Education, Department of Education, National Chiayi University, Taiwan // [2]Chiayi County Jhuci Senior High School, Taiwan // [3]Taipei Municipal Linong Elementary School, Taiwan // chuck@mail.ncyu.edu.tw // willy76616@gmail.com // tedbob51@gmail.com

[*]Corresponding author

ABSTRACT: During the worldwide pandemic of coronavirus disease 2019 (COVID-19 pandemic), online learning is increasingly vital for students to learn at home, and online learning platforms provide learning opportunities to students. The Junyi Academy online platform is an online learning platform that both helps lower-achieving students review lessons and helps teachers in Taiwan do differentiated instruction. Several studies have shown the relationships between students' attention and their academic achievements for students' self-learning, but how to best use these platforms to help students learn by themselves is unclear. Therefore, this study investigates the relationships between students' attention and their academic achievements with two online learning environments. A total of 38 upper secondary students in Taiwan to participate in this study, and these students were divided between a Khan-style video lecture (VL) group and an online practice (OP) group. This study adopted an experimental design with data collected by an electroencephalogram (EEG). The results show that students' attention in the VL group was higher than in the OP group. Furthermore, their attention in three stages differed between the two groups. Student attention was similar in the two groups for the first stage, but the VL group had higher attention for the second and third stages was than did the OP group. In addition, there was no relationship between students' attention and their academic achievements in the VL and OP groups. Finally, this study raised some suggestions the future research.

Keywords: Attention, Video lecture, Online practice, Online learning platform, Electroencephalogram (EEG)

1. Introduction

The Programme for International Student Assessment (PISA) has analysed the relationships between students' socio-economic status and their mathematics achievements since 2012, finding that students' achievements in some countries, including Taiwan, could be predicted by their socio-economic status (SES) (OECD, 2014). To reduce the gap between high- and low achieving students, Taiwanese government promotes some programs, such the Project for the Implementation of Remedial Instruction and the Educational Priority Area program, to enhance low-achieving student achievements, Sung et al. (2014) adopted a time series analysis to determine whether participating in these programs could improve student achievements. This study found that the achievement gap in Taiwan increased between higher- and lower-SES. Both results in PISA from 2012 (OECD, 2014) and Sung et al. (2014) showed a similar trend between student achievements and their parents' SES. Therefore, the Taiwanese government has expended much effort to reduce the gap between high- and low achieving students, and OECD (2018) indicated that the gap between students' achievements and their parents' SES has been reduced in Taiwan. However, SES is still a strong predictor to explain 13% of the variation in Taiwanese students' mathematics achievements (OECD, 2020). This may be because higher-SES students have better educational resources. Hwang (2015) pointed out that many secondary students in Taiwan go to cram schools after school, but many lower-SES students' parents could not afford the extra tuition fee. Thus, online learning materials on e-learning platforms, such as the Junyi Academy online platform, are useful to help lower-SES students review lessons, and schools in Taiwan also provide relevant hardware, such as tablets, to help these students review lessons at home after school.

After the global COVID-19 pandemic began in early 2020, the World Health Organization announced that people should maintain social distancing and wear a mask both indoors and outdoors (World Health Organization, n.d.). In addition, schools have closed to limit the spread of this disease in numerous countries, limiting students' learning opportunities, so many schools in these countries use synchronous online teaching to help students learn at home (Bailey et al., 2020; Jan, 2020). Although there were only 799 confirmed COVID-19 cases in Taiwan by the end of 2020, Taiwan's government still encouraged each school to simulate school closures and demonstrate how to teach/learn at home (Ministry of Education, n.d.). Taiwan's Ministry of Education (n.d.) suggests several ways to teach/learn at home, including using Google Meet, Microsoft Teams, and several e-learning platforms. The Junyi Academy online platform is one of e-learning platforms and aims to help educators teach students in accordance with their aptitude and to motivate their interest in learning (Junyi Academy, n.d.). Under COVID-19 restrictions, this online platform is used to help primary and secondary students learn at home, but how to use these platforms appropriately to enhance student achievement is still unclear to many teachers.

Students' academic achievements is affected by several factors, one of which is their attention since it can determine academic achievement (Bester & Brand, 2013; Chang et al., 2019; Shadiev et al., 2017). Attention is defined as the limitations in processing information and how people monitor these limitations. The five types of attention are focusing, perceptual enhancement, binding, sustaining behaviour, and action selection (Medin et al., 2005). Each type of attention explains a different aspect of attention, and this study considers attention as sustained attention, which is when someone pays attention to salient objects and excludes other objects for a certain amount of time (Medin et al., 2005). In online learning environments, students may have interference from other icons or pop-up messages, and their sustained attention would indicate their concentration on learning. Online learning environments have pros and cons for student learning. On one hand, Terras and Ramsay (2015) indicate that students are able to watch online video lectures repeatedly to increase learning opportunities. On the other hand, Lodge and Harrison (2019) find that using technology to learn has a negative impact on the brain. In addition, Lin and Chen (2019) show that students using online video lectures to learn might ignore important content, and Chen et al. (2017) state that students would be distracted without teacher supervision in online learning environments. However, using online learning environments is indispensable during the current COVID-19 pandemic. Therefore, this study investigates relationships between students' attention and their academic achievements in different online learning environments. Accordingly, three research questions are addressed in the next section.

1.1. Research questions

(1) What were differences between students' pre-test and post-test scores under two types of online learning environments?
(2) What were differences in students' attention at different learning stages in two types of online learning environments?
(3) What were relationships between students' attention and their academic achievement in two types of online learning environments?

2. Theoretical frameworks

This section introduces relevant theories to analyse two types of online learning environments, video lecture and online practice, as used in the Junyi Academy online platform. Then empirical studies examine relationships between students' academic achievements and their attention in online environments.

There are two main theories to describe students' multimedia learning. One is the cognitive theory of multimedia learning (CTML, Mayer, 2014) and the other one is cognitive load theory (CLT, Paas & Sweller, 2014). Both theories are related to the limitations of working memory and the capacity of working memory determines how students select information, and information selection is associated with sustained attention. CTLM explains interactions between pictures and text on students' learning (Mayer, 2014), whereas CLT outlines three categories of instructional design to reduce students' working memory load, including extraneous, intrinsic and germane cognitive load (Paas & Sweller, 2014). Different online learning environments could be based on different theories. In this study, two types of online learning environments are considered in the Junyi Academy online platform. One type is "video lecture (VL)," and the other type is "online practice (OP)." The type of video lecture in Junyi Academy online platform is referred to as a "Khan-style video lecture," and Chen and Wu (2015) define "the Khan-style video lecture" as a handwritten tutorial with digital pens and tablets with an audio voice to explain content. Students would not see a lecturer's face and no gestures guide students to see what content is important in this online learning environment. Students have to pay attention to hear what lecturers say and follow the lecturers' voice to learn. For online practice, the Junyi Academy online platform posts a problem for students to solve once, and there is a yellow icon to provide hints to help solve problems. Students can solve problems on real paper and submit an answer online in this environment.

According to both online learning environments, the Khan-style video lecture is related to a teaching method and could be explained by the extraneous cognitive load. The extraneous cognitive load represents that information is given by instructional designers, and the way of giving information is related to teaching activities (Chandler & Sweller, 1991). This means that lecturers in the Khan-style video lecture environment have selected relevant information to demonstrate to students. However, students in the Khan-style video lecture environment have to use visual and auditory channels to learn, and Ayres and Sweller (2014) indicate that using both channels to learn would increase students' cognitive load due to the need to coordinate information from different channels. This implies that students might have to pay attention closely to learn in the Khan-style video lecture, and thus spend more cognitive load to learn (Chen & Wu, 2015). Online practice is different from the Khan-style video lecture since an interface in the online practice learning environment only shows a problem with yellow and green icons (see Figure 2). When students click the yellow icon, a hint is shown by Arabic numerals in yellow, so students solve problems easily. The design of the online practice learning environment followed the signalling principle in CTML (Mayer, 2014). Glaser et al. (2017) claim that signalling can help students organize information. The yellow icon with relevant solutions guides students to organize information in order to solve problems. Therefore, the yellow icons with relevant hints are signals to highlight relevant points in an online practice environment. Kalyuga et al. (1999) claim that using signalling to design an online environment could reduce the load on students' working memory, and this study hypothesises that students in an online practice environments might require less cognitive load to learn. In sum, students have more cognitive load in the Khan-style video lecture environments, whereas students in the online practice environments could use less cognitive load to learn effectively.

2.1. Students' attention and their academic achievements

As in the theories of multimedia learning discussed above, cognitive load is related to sustained attention, and there are three approaches to evaluate people's sustained attention in online learning environments, including an electroencephalogram (EEG), eye tracking and paper-and-pencil tests. Detecting people's sustained attention generally uses an EEG test with NeuroSky's MindWave earphone, and the EEG signal data from the NeuroSky has been validated (Chen & Huang, 2014; Chen & Wang, 2018; Chen et al., 2017; Chen & Wu, 2015; Shadiev et al., 2017; Sun & Yeh, 2017; Wu et al., 2020). Numerical data from the EEG signals reflects in real time how many nervous system activities are related to people's sustained attention (Chen et al., 2017). Higher values from the EEG indicate higher attention. However, relationships between attention and academic achievements show inconsistent conclusions in different learning environments. The relationships between students' academic achievements and their attention are discussed below to show which factors might influence these relationships:

Students using online learning environments could have higher academic achievement and higher-level attention than those in traditional environments. Chang et al. (2019) compared a traditional PowerPoint lecture with the massive online courses (MOOCs) in lower secondary students' attention and their academic achievements, and this study found that students using MOOCs could have higher achievement and higher-level attention. Shadiev et al. (2017) used technology to teach English, finding that university students using technology had higher attention by EEG and had better learning performances in comparison with students without using technology to learn English. Although online learning environments have advantages for learning, they need extra support to help students concentrate on learning. Chen and Wang (2018) indicated that students who had extra support (monitoring and alarm mechanisms) in online environments would get more attention and have better academic achievement.

However, more attention might not lead to better academic achievement because of students' cognitive load. Although Chen and Wu (2015) showed that university students' performed similarly after different online learning environments, the current study finds that students viewing slides with a lecturer's voice and image had to use more cognitive load than students viewing slides without a lecturer's voice. Furthermore, Wu et al. (2020) compared digital game-based learning environments (DGLE) and static E-learning environments (SELE), and found that university students' attention and achievement in the two groups were similar. Wu et al. (2020) indicated that students were interested in DGLE and DGLE could trigger students' learning motivation, but DGLE was more complex and increased students' cognitive load. A complex learning environment might overload students' cognitive capacity. Both Chen and Wu (2015) and Wu et al. (2020) imply that more complex online environments lead to higher attention, though higher attention does not necessarily bring better academic achievements.

The relationships between students' attention and their academic achievements are still debated, but these relationships might be associated with student age. Chen and Wang (2018) and Sun and Yeh (2017) developed attention-monitoring systems for online learning environments. While students in Chen and Wang (2018) were lower secondary students, Sun and Yeh (2017) used university students. Both studies found that students experiencing attention monitoring systems had higher attention, but students' achievements showed different patterns. The different patterns could be explained by differences in cognitive executive functions. Youths' cognitive executive functions are still developing before 13 years old (Davidson et al., 2006). Although students in Sun and Yeh's (2017) study without attention monitoring systems had lower attention, these students could rely on their relevant prior knowledge and working memory to do a post-test. Their cognitive executive functions are mature and could overcome their lower attention to get acceptable scores in the post-test. In Chen and Wang (2018), primary students' cognitive executive functions were still developing, so their prior knowledge and working memory were limited. This might explain why primary students with higher attention could get better academic achievement, while mature students could be supported by their prior knowledge and relevant experiences in an achievement test.

According to the aim of the Junyi Academy online platform mentioned above, this platform has a Khan-style video lecture and online practice for each student, to choose and their cognitive load should be different in using the two online environments. This study reveals some potential research values. Some studies (Chang et al., 2019; Chen & Wang, 2018; Chen et al, 2017; Chen & Wu, 2015; Lin & Chen, 2019) used an EEG to detect sustained attention in video lecture online environments, and most studies (Chang et al., 2019; Chen & Wang, 2018; Chen et al., 2017; Chen & Wu, 2015) analysed an effect of instructional immediacy. However, only Lin and Chen (2019) discussed the learning effect after reviewing, using EEG to detect 55 primary students' attention and their achievements after reviewing. That study used an attention-based video lecture review mechanism (AVLRM) to evaluate student achievement and found that only low-attention students using AVLRM had higher achievement than without using AVLRM. In addition, students' attention in online practice environments is evaluated by eye-tracking (Glaser et al., 2017). Overall, few studies have evaluated the learning effect after reviewing, and no study has used the EEG to find effects of the signalling principle. This study provides relevant empirical data to expand the theoretical basis of multimedia learning techniques.

3. Research methods

This section describes participants, materials, procedures and data analysis in order to fulfil the goals of this study and research questions. It also demonstrates how this study answers these research questions.

3.1. Participants

All participants in this study were from the same regional secondary school in Chiayi County, Taiwan. Fifty Taiwanese secondary students were recruited for this study. This study also informed consent from the students and their families. Students with pre-test scores higher than 17 points were excluded, leaving thirty-eight participants for the study. Twenty-four grade 11 students and fourteen grade 12 students were randomly assigned the video lecture (VL) and the online practice (OP) groups. Each group had 19 students. In the pre-test, students' scores had no significant difference by both types of online learning environments and grades, online learning environments: $t(36) = .17$, $p = .839$, Cohen's $d = .066$; grade: $t(19.50) = 1.49$, $p = .154$, Cohen's $d = .664$. Thus students' pre-test scores for the two types of online learning environments and grades were similar. Students' pre-attention had also no significant difference for the two types of online learning environments and grades, online learning environments: $t(36) = 1.13$, $p = .267$, Cohen's $d = .366$; grade: $t(36) = .17$, $p = .863$, Cohen's $d = .058$. Thus students' attention spans before reviewing lessons online were similar. Students' pre-test scores and pre-attention are shown in Table 1.

Table 1. Students' pre-test scores and pre-attention in the two types of online learning environments and grades

Types	Video Lectures (VL)				Online Practice (OP)			
Grade	Grade 11		Grade 12		Grade 11		Grade 12	
Number	11		8		13		6	
	Test	Attention	Test	Attention	Test	Attention	Test	Attention
M	16.09	41.70	12.75	39.07	14.46	44.80	15.50	47.23
SD	1.70	15.68	4.10	15.43	2.88	12.07	3.62	11.93

3.2. Materials

The midsegment theorem was the main topic in this study, using the VL and OP groups to present this theorem. The topic of the midsegment theorem was introduced in grade 9, and students apply this theorem for trigonometric functions in grade 11 (Ministry of Education, 2014). Although upper secondary students should understand the midsegment theorem, grade 11 students must review this theorem to help them learn the trigonometric functions, and grade 12 students could review it for the university entrance exam. In this study, the midsegment theorem was for reviewing in grades 11 and 12 students. In the VL group, a lecturer first introduces the midsegment theorem first and then demonstrates two problems to solve. Students in the OP group receive 10 to 20 problems to solve. Students who could not finish any of these problems could click an icon to show how to solve problems. For students who still could not solve these problems, their computer screen would have pop-up part to show a hint. After a problem is solved correctly, the online system shows the next problem. Materials in both the VL and OP group were from the Junyi Academy online platform. The interface samples of the VL and OP groups are shown in Figure 1 and 2, respectively.

Figure 1. Presentation in the VL group

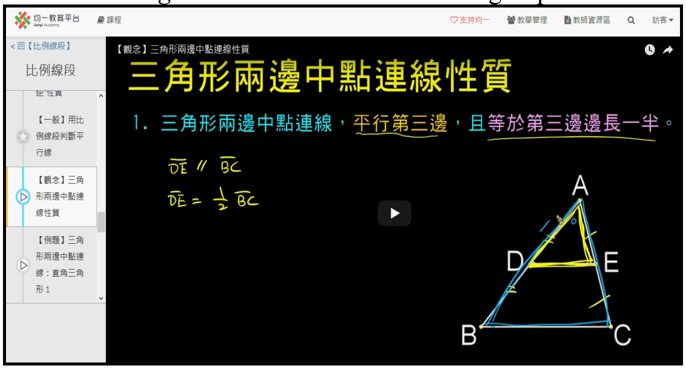
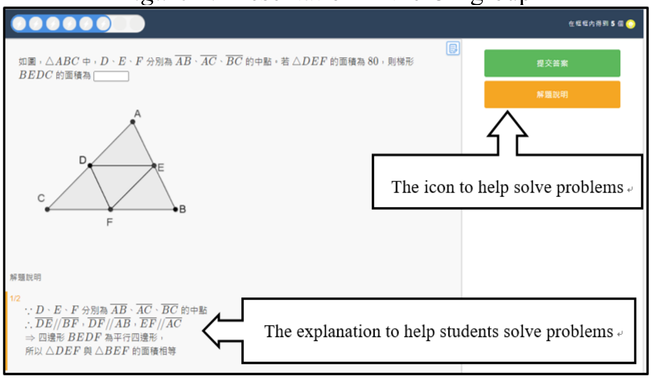


Figure 2. Presentation in the OP group



To understand students' academic achievement, this study adapted an achievement test from the Junyi Academy online platform. The test had 21 items, and all items' item discrimination index (D value) should be higher than 0.4. According to this criterion, this study did the pilot study from 52 students' responses. The result showed that there were two items which should be excluded because their D values were lower than 0.4. Therefore, only nineteen items were in the achievement test, with D values ranging from 0.46 to 0.92. Each item had 1 point, so the full score in this test was 19 points. Validity was verified by experienced high school mathematics teachers, and the reliability was .86. It was used as both a pre- and post-test.

Physiological signals were used to evaluate students' levels of attention. The instrument, as shown in Figure 3, is a brainwave sensing headset to act as an electroencephalogram (EEG) from NeuroSky technologies. Students wear the headset with a sensor and the sensor receives students' brainwaves and transmits them to a computer through Bluetooth. Output values are from 0 to 100, with higher values indicating higher levels of attention.

Figure 3. Brainwave sensing headset (NeuroSky, n.d.)

3.3. Procedures

All students took the achievement pre-test with no time limit before online learning. After one week, groups of from 1 to 3 students did online learning in their school's computer lab. In order to understand the baseline of students' attention (pre-attention), all students would wear the headset and close their eyes to relax for 3 minutes. Examiners asked students to open their eyes and start to review the midsegment theorem online within eight minutes. After reviewing online, the students took the same achievement test as a post-test. Instructions for the VL and OP groups were as follows: (1) For the VL group's students: "This experiment has three stages. The first stage is to help review what the midsegment theorem is, and the second and third stages each have one problem to solve. You have eight minutes to review online, and you should move the screen cursor to check your learning progress." After reviewing online, the examiners would ask students do the post-test. (2) For the OP group's students: "This experiment has three stages, and each stage has different problems to solve. You have eight minutes to solve problems. If you do not know how to solve these problems, you can click a yellow icon to get some hints.

3.4. Data analysis

This study used an experimental design to understand the effects of different types of online learning environments on students' attention and academic achievement. As three research questions were addressed, this study used repeated measure ANOVA for research questions 1 and 2, and correlation analysis for research question 3. For research question 1, this study used students' pre- and post-test scores in the achievement test to compare the two types of online learning environments by analysing the repeated measure ANOVA. The types of online learning environments and the pre- and post-test were independent variables, and scores in the pre- and post-test were dependent variables. For question 2, this study divided the experiments into three sections, with students in different groups having different sections. For the VL groups, introducing the midsegment theorem was the first section, and demonstrating two problems were the second and third sections. The cut points in the two sections were 3 minutes 25 seconds, 5 minutes to record students' attention. The first stage was 0 to 3 minutes 25 seconds, the second stage is from 3 minutes 25 seconds to 5 minutes, and the third stage is longer than 5 minutes. For students in the OP group, the cut points were the same as the VL group. The recording students' attention was recorded in seconds. The types of online learning environments and the three stages were independent variables, and the value of students' attention in three stages was the dependent variable. This research question also used repeated measure ANOVA to show the trend of students' attention in three stages for the two types of online learning environments. According to the variables in research questions 1 and 2, the data of student scores in the achievement test and the value of attention were numerical data, while the different types of learning environments and the pre- and post-test were categorical data. For research question 3, this study used correlation analysis to determine the relationships between student achievement and attention in the two types of online learning environments. Students' post-test scores were their achievement and students' overall attention from the EEG data was their attention. Both student achievement and attention were numerical data. Correlation analysis showed whether students' attention had a positive effect on student achievement

4. Results

Statistical analysis was used to answer the research questions, and this section has three parts to consider each question

4.1. Research question 1

Two-way repeated measure ANOVA was used to answer research question 1. Box's test for equivalence of covariance matrices showed no significant difference, with Box's $M = 5.25$, $F(3, 233280) = 1.64$, $p = .177$. Thus variances in students' pre- and post-test scores were similar, and the data could use the two-way repeated measure ANOVA to test. Results showed that students' scores for pre- and post-test had a significant difference, $F(1, 36) = 11.60$, $p = .002$, $\eta^2 = .244$, and student scores in the post-test were higher than their pre-test. However, there was no significant difference between two types of online learning environments, $F(1, 36) = .004$, $p = .949$, $\eta^2 = .000$. Meanwhile, there was no interaction effect between student scores in the pre- and post-test and different types of online learning environments, $F(1, 36) = .35$, $p = .561$, $\eta^2 = .009$. This indicates that students' scores between the pre- and post-test in the VL and OP groups were similar and students' scores between the pre- and post-test were not influenced by different online learning environments. Descriptive statistics are shown in Table 2.

Table 2. Descriptive statistics for students' pre-and post-test scores with the two types of online learning environments

Types	Video Lectures (VL)		Online Practice (OP)	
Test	Pre-test	Post-test	Pre-test	Post-test
M	14.68	15.95	14.47	16.62
SD	3.32	2.70	3.03	2.38

4.2. Research question 2

To answer research question 2, this study used the same statistical method as for research question 1. Box's test for equivalence of covariance matrices showed no significant difference, with Box's $M = 7.80$, $F(6, 9389.90) = 1.18$, $p = .313$. Thus the variances in students' attention were similar, and the data could use the two-way repeated measure ANOVA. The results showed that students' attention in all three stages had no significant difference, $F(2, 72) = .60$, $p = .554$, $\eta^2 = .016$, but students' attention between different online learning environments had a significant difference, $F(1, 36) = 13.50$, $p = .001$, $\eta^2 = .273$. Meanwhile, students' attention in three stages and using different online learning environments showed an interaction effect, $F(2, 72) = 4.26$, $p = .018$, $\eta^2 = .106$. Therefore, the simple main effect was needed to understand which factor to determine the interaction effect.

This study used a t-test and ANOVA to test the simple main effect of students' attention between each stage and two types of online learning environments. For the VL group's attention, student attention between three stages had no significant difference, $F(2, 36) = 1.02$, $p = .371$, $\eta^2 = .054$. Thus students' attention in the VL group between three stages was similar. For the OP group' attention, students' attention between three stages had a significant difference, $F(2, 36) = 5.40$, $p = .009$, $\eta^2 = .231$. This showed that students' attention in the OP group between three stages was different, and post hoc analysis was used for the OP group. Student attention in the first stage was better than in the second and third stages, but there was no difference in attention between the second and third stages. Thus the OP group students in the first stage could pay more attention than in the second and third stages. The results indicated that students in the first stage were more attentive than in the second and third stage, but their attention levels in the second and third stages were similar for the OP group's students. Thus students' attention in the VL group was similar between three stages, but students' attention in the OP group had different patterns since students were more attentive in the first stage, and less so in the second and third stages.

Students' attention in the second and third stages had significant differences; second: $t(36) = 3.58$, $p = .001$, Cohen's $d = 1.163$; third: $t(36) = 4.72$, $p < .001$, Cohen's $d = 1.531$. In the second stage, students' attention in the VL group was higher than in the OP group. Students' attention in the third stage showed the same pattern. Thus students in the VL group were more attentive than those in the OP group in the second and third stages. However, there was no significant difference in student attention for the first stage between different online learning environments, $t(36) = .87$, $p = .393$, Cohen's $d = .280$. This indicates that students' levels of attention in the first stage were similar for the two types of online learning environments.

According to the results in research question 2, students' attention levels were similar for the VL and OP group in the first stage, and students' attention levels in the three stages were similar in the VL group. For the OP group, students' attention levels decreased in the second and third stages from the first stage. Descriptive statistics are shown in Table 3.

Table 3. Descriptive statistics of student attention in three stages in two types of online learning environments

Types	Video Lectures (VL)			Online Practice (OP)		
Stage	First	Second	Third	First	Second	Third
M	53.95	55.37	56.71	51.83	47.90	48.20
SD	8.11	6.91	6.43	6.97	5.90	6.99

4.3. Research question 3

Table 4 shows descriptive statistics for student achievement and their attention. However, there was no significant correlation between student achievements and their attention in the VL and OP groups, VL: $r(17) =$

.22, p = .364; OP: $r(17)$ = -.14, p = .571. Although the relationships between student achievement and their attention in the VL and OP groups showed different trends, the relationships were irrelevant.

Table 4. Descriptive statistics of student achievement and attention in two types of online learning environments

Types	Video Lectures (VL)		Online Practice (OP)	
Variable	Achievements	Overall Attention	Achievements	Overall Attention
M	15.95	55.37	16.26	49.34
SD	2.70	5.32	2.38	4.90

5. Discussion

This section addresses the research questions to answer with relevant theories and empirical data for discussion

5.1. Students achievement between two types of online learning environments

This study found similar student achievement levels in the VL and OP groups, in contrast to previous studies (Bester & Brand, 2013; Chang et al., 2019; Chen & Wang, 2018; Lin & Chen, 2019; Shadiev et al., 2017). This might be due to differences in learning environments. These studies discussed students' achievements between the online and traditional learning environments or in using the same online learning environment with/without extra supports. However, this study compared two different online environments with the same topic, and the student achievement results in this study are supported by the studies of Chen and Wu (2015), Hew and Lo (2020) and Ilioudi et al. (2013).

These studies (Chen & Wu, 2015; Hew & Lo, 2020; Ilioudi et al., 2013) compared Khan-style video lectures with other types of video lectures to evaluate students' academic achievement. Chen and Wu (2015) indicated that three types of video lectures in their study could improve students' achievements, but students using the Khan-style video lectures did not improve as much as those using recording classroom lectures or the lecturer's image with lecture slides. Ilioudi et al. (2013) compared students' achievements between recorded classroom lectures, Khan-style lectures and a printed book, with results similar to Chen and Wu (2015). Chen and Wu (2015) claimed that the Khan-style video lecture had no better visual layouts to guide students to learn in online environments, and the inappropriate layouts might influence students' learning performance. In addition, Hew and Lo (2020) stated that using Khan-style video lectures with teacher's talking head videos would increase students' achievement scores. Hew and Lo (2020) implied that the Khan-style video lecture should be used as supplementary material to help students review lessons. Ilioudi et al. (2013) also noted that the Khan-style video lecture was not appropriate for students' self-learning because they had no interactive opportunity to ask questions. Thus, the VL online environment could help students review the midsegment theorem, but this online learning environment might have restricted students' learning in some conditions

The OP learning environment could improve students' learning achievement, which could be due to the repetitive practice of traditional learning methods in Taiwan: (Yang & Lin, 2015). The OP learning environments provide several questions to help students review the midsegment theorem, and Taiwanese students also use a similar method to learn mathematics. Icons in the OP learning environment of the Junyi Academy online platform are like scaffolding to help students how to solve mathematical problems. By doing more practice, students have higher achievement. Although this traditional learning method may not be the best way to learn mathematics, it is effective (Mullis et al., 2012). According to the results for research question 1, the VL and OP online environments in this study could both help students recall what they learned, and students could do practice problems and check their answers. This could explain why both online environments could improve students' academic achievement, but the different interfaces in the VL and OP groups might explain student achievements in the two groups.

5.2. Students' attention in the two types of online learning environments

Although student achievement levels in the VL and OP learning environments were similar, students' attention in the two learning environments showed different trends. Overall, students' attention in the VL group was higher than in the OP group, and students' level of attention differed between the second and third stages. Students in the VL group would see a lecturer to explain the midsegment theorem with the lecturer's handwriting and voice only. As stated above, students could have increased their cognitive load for a higher level of attention, as shown by the EEG data. Chen and Wu (2015) state that students using a Khan-style video lecture would have their attention distracted in order to integrate information, thereby inhibiting their learning. Although students in the VL group showed a higher attention, this might be because they needed to focus on integrating relevant information while watching this video lecture. Students in the VL group need to process several pieces of information at the same time, which would increase their load on working memory. Students' attention level reflects how they control visual information in their working memory (Lodge & Harrison, 2019). Students in the VL group had to process visual and auditory information at the same time through dual channels (Mayer, 2014; X. Yang et al., 2020).Taken together, these studies (Chen & Wu, 2015; Mayer, 2014; X. Yang et al., 2020) indicate that students in the VL group needed to follow a lecturer's voice to track his/her handwriting and see graphic images to understand the midsegment theorem. Alpizar et al. (2020) used meta-analysis to show that the online environment with relevant images and texts would increase students' cognitive load. This means that students in the VL online environment would have increased cognitive load due to integrating the voice, handwriting and images. This increased cognitive load might have caused their higher attention.

The interface in the OP learning environment might be designed according to the signalling principle, which would reduce students' cognitive load (Mayer, 2014). Alpizar et al. (2020) also found that using computers together with printed information was a better online learning environment. Students in the OP group would do arithmetic to review the midsegment theorem, and these students needed to calculate on paper and then submit an answer on the platform. Working on paper without mental calculation might reduce students' cognitive load to influence their attention. However, Ilioudi et al. (2013) indicated that students in using technology would initially spend additional time to be familiar with an interface in order to learn it. This could explain why the OP group's students had higher attention in the first stage because they were familiar with an interface in this learning platform. Accordingly, students could use less attention to learn in the OP learning environments.

5.3. Students achievement and attention in the two types of online learning environments

The results for research question 3 differ from previous studies (Chen & Huang, 2014; Chen & Wang, 2018; Lin & Chen, 2019), which might be due to the developmental level of students' cognitive executive functions. Participants in these studies were primary or lower secondary students, whose achievement is still determined by their attention. Students might have learned relevant topics before, and then use their prior knowledge to solve problems in this study. This possibility is supported by Sun and Yeh's (2017), who indicated that whether students' relevant prior knowledge/experience would affect the relationships between their attention and achievement. Students in this study had learned the midsegment theorem, so their prior knowledge would influence the relationships between attention and academic achievements.

6. Conclusions, limitations and future research

This study found that students in the VL group had higher attention when learning, but students in the OP group could use less attention to learn. Both learning environments could enhance student achievement. Although both environments in this study helped students review lessons, this study reveals their pros and cons for learning. The relationships between student achievement and their attention might be influenced by their cognitive executive functions and their prior knowledge. Following the pandemic, educators should consider how to use online environments to help students learn outside of the classroom.

There were some limitations in this study. First of all, there were only 38 participants, so it would be difficult to extend this study's results to other students. Secondly, this study used an existed online learning platform, with limited learning environments, so other variables could not be manipulated for discussion. Finally, this study asked students to finish reviewing within a limited time restriction, which might affect student performance. Accordingly, future research could develop or modify current interfaces to more clearly analyse the relationships between student achievement and attention and help students learn better by themselves. This study used the EEG to detect students' attention in the OP learning environment, and future research could combine the EEG with eye tracking to more closely monitor students' attention.

Acknowledgement

This study is supported in part by the Ministry of Science and Technology of Taiwan under Contract Numbers MOST 108-2511-H-415-005, MOST 109-2511-H-415-004 and MOST 110-2511-H-415-002.

References

Alpizar, D., Adesope, O. O., & Wong, R. M. (2020). A Meta-analysis of signaling principle in multimedia learning environments. *Educational Technology Research and Development, 68*(5), 2095-2119. https://doi.org/10.1007/s11423-020-09748-7

Ayres, P., & Sweller, J. (2014). The Split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 206-226). Cambridge University Press.

Bailey, D., Almusharraf, N., & Hatcher, R. (2020). Finding satisfaction: Intrinsic motivation for synchronous and asynchronous communication in the online language learning context. *Education and Information Technologies, 26,* 2563–2583. https://doi.org/10.1007/s10639-020-10369-z

Bester, G., & Brand, L. (2013). The Effect of technology on learner attention and achievement in the classroom. *South African Journal of Education, 33*(2), 1-15. https://doi.org/10.15700/saje.v33n2a405

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293-332. https://doi.org/10.1207/s1532690xci0804_2

Chang, J.-J., Lin, W.-S., & Chen, H.-R. (2019). How attention level and cognitive style affect learning in a MOOC environment? Based on the perspective of brainwave analysis. *Computers in Human Behavior, 100*, 209-217. https://doi.org/10.1016/j.chb.2018.08.016

Chen, C.-M., & Huang, S.-H. (2014). Web-based reading annotation system with an attention-based self-regulated learning mechanism for promoting reading performance. *British Journal of Educational Technology, 45*(5), 959-980. https://doi.org/10.1111/bjet.12119

Chen, C.-M., & Wang, J.-Y. (2018). Effects of online synchronous instruction with an attention monitoring and alarm mechanism on sustained attention and learning performance. *Interactive Learning Environments, 26*(4), 427-443. https://doi.org/10.1080/10494820.2017.1341938

Chen, C.-M., Wang, J.-Y., & Yu, C.-M. (2017). Assessing the attention levels of students by using a novel attention aware system based on brainwave signals. *British Journal of Educational Technology, 48*(2), 348-369. https://doi.org/10.1111/bjet.12359

Chen, C.-M., & Wu, C.-H. (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education, 80*, 108-121. https://doi.org/10.1016/j.compedu.2014.08.015

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*(11), 2037-2078. https://doi.org/10.1016/j.neuropsychologia.2006.02.006

Glaser, M., Lengyel, D., Toulouse, C., & Schwan, S. (2017). Designing computer-based learning contents: Influence of digital zoom on attention. *Educational Technology Research & Development, 65*(5), 1135-1151. https://doi.org/10.1007/s11423-016-9495-9

Hew, K. F., & Lo, C. K. (2020). Comparing video styles and study strategies during video-recorded lectures: Effects on secondary school mathematics students' preference and learning. *Interactive Learning Environments, 28*(7), 847-864. https://doi.org/10.1080/10494820.2018.1545671

Hwang, T.-M. (2015). The Studying and striving of secondary students. In S. Hsu & Y.-Y. Wu (Eds.), *Education as Cultivation in Chinese Culture* (pp. 127-148). Springer. https://doi.org/10.1007/978-981-287-224-1_7

Ilioudi, C., Giannakos, M., & Chorianopoulos, K. (2013). Investigating differences among the commonly used video lecture styles. *Proceedings of the WAVe 2013: The Workshop on Analytics on Video-based Learning* (pp. 21-26). CEUR. http://ceur-ws.org/Vol-983/WAVe2013-Proceedings.pdf

Jan, A. (2020). A Phenomenological study of synchronous teaching during COVID-19: A Case of an international school in Malaysia. *Social Sciences & Humanities Open, 2*(1), 100084. https://doi.org/10.1016/j.ssaho.2020.100084

Junyi Academy. (n.d.). *均一教育平台：關於我們 [Junyi Academy online platform: About us].* https://official.junyiacademy.org/about/

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*(4), 351-371. https://doi.org/10.1002/(SICI)1099-0720(199908)13:4<351::AID-ACP589>3.0.CO;2-6

Lin, Y.-T., & Chen, C.-M. (2019). Improving effectiveness of learners' review of video lectures by using an attention-based video lecture review mechanism based on brainwave signals. *Interactive Learning Environments, 27*(1), 86-102. https://doi.org/10.1080/10494820.2018.1451899

Lodge, J. M., & Harrison, W. J. (2019). The role of attention in learning in the digital age. *Yale Journal of Biology and Medicine, 92*(1), 21-28.

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Eds.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 43-71). Cambridge University Press.

Medin, D. L., Ross, B. H., & Markman, A. B. (2005). *Cognitive psychology* (4th ed.). John Wiley & Sons.

Ministry of Education. (n.d.). *教育雲* [*Learning cloud]*. Ministry of Education. http://learning.cloud.edu.tw/onlinelearning/#

Ministry of Education. (2014). *12-year compulsory education curriculum guideline*. National Academy for Educational Research. https://www.naer.edu.tw/ezfiles/0/1000/attach/87/pta_5320_2729842_56626.pdf

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. TIMSS & PIRLS International Study Center, Boston College.

NeuroSky. (n.d.). *Brainwave Sensing Headset*. NeuroSky. https://store.neurosky.com/pages/mindwave

OECD. (2014). *PISA 2012 results: What students know and can do – Student performance in mathematics, reading and science* (Revised ed. Vol. I). PISA, OCED Publishing. https://doi.org/10.1787/9789264208780-en

OECD. (2018). *Equity in education: Breaking down barriers to social mobility*. PISA, OECD Publishing. https://doi.org/10.1787/9789264073234-en.

OECD. (2020). Students' socio-economic status and performance. In *PISA 2018 Results (Volume II)*: *Where All Students Can Succeed* (pp. 49-62). PISA, OECD Publishing. https://doi.org/10.1787/f7986824-en

Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2 ed., pp. 27-42). Cambridge University Press.

Shadiev, R., Huang, Y.-M., & Hwang, J.-P. (2017). Investigating the effectiveness of speech-to-text recognition applications on learning performance, attention, and meditation. *Educational Technology Research & Development, 65*(5), 1239-1261. https://doi.org/10.1007/s11423-017-9516-3

Sun, J. C.-Y., & Yeh, K. P.-C. (2017). The Effects of attention monitoring with EEG biofeedback on university students' attention and self-efficacy: The Case of anti-phishing instructional materials. *Computers & Education, 106*, 73-82. https://doi.org/10.1016/j.compedu.2016.12.003

Sung, Y.-T., Tseng, F.-L., Kuo, N.-P., Chang, T.-Y., & Chiou, J.-M. (2014). Evaluating the effects of programs for reducing achievement gaps: A Case study in Taiwan. *Asia Pacific Education Review, 15*(1), 99-113. https://doi.org/10.1007/s12564-013-9304-7

Terras, M. M., & Ramsay, J. (2015). Massive open online courses (MOOCs): Insights and challenges from a psychological perspective. *British Journal of Educational Technology, 46*(3), 472-487. https://doi.org/10.1111/bjet.12274

World Health Organization. (n.d.). *Coronavirus disease (COVID-19) advice for the public*. World Health Organization. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public

Wu, C.-H., Tzeng, Y.-L., & Huang, Y.-M. (2020). Measuring performance in leaning process of digital game-based learning and static E-learning. *Educational Technology Research and Development, 68*(5), 2215-2237. https://doi.org/10.1007/s11423-020-09765-6

Yang, K.-L., & Lin, F.-L. (2015). The effects of PISA in Taiwan: Contemporary assessment reform. In K. Stacey & R. Turner (Eds.), *Assessing Mathematical Literacy: The PISA Experience* (pp. 261-273). Springer International Publishing. https://doi.org/10.1007/978-3-319-10121-7_14

Yang, X., Lin, L., Wen, Y., Cheng, P.-Y., Yang, X., & An, Y. (2020). Time-compressed audio on attention, meditation, cognitive load, and learning. *Educational Technology & Society, 23*(3), 16-26.

Sun, Y., Chang, C.-H., & Chiang, F.-K. (2022). When Life Science Meets Educational Robotics: A Study of Students' Problem Solving Process in a Primary School. *Educational Technology & Society*, *25* (1), 166-178.

When Life Science Meets Educational Robotics: A Study of Students' Problem Solving Process in a Primary School

Yulong Sun[1], Chun-Hao Chang[2] and Feng-Kuang Chiang[1*]
[1]Department of Educational Technology, College of Education, Shanghai Normal University, China //
[2]Department of Allied Health Education and Digital Learning, National Taipei University of Nursing and Health Sciences, Taiwan // sunyulong32@163.com // chunhao@gm.ntunhs.edu.tw // fkchiang@shnu.edu.cn
[*]Corresponding author

ABSTRACT: Previous studies on the topic of "Problem solving" indicate it's one of the skills of students in the 21st century that educational robots can effectively support. Additionally, there are gender differences in the problem-solving process. Understanding the problem-solving process and using knowledge to solve problems is key to improving one's problem-solving ability. We therefore conducted a study with 69 fifth graders aimed at exploring whether educational robots can help students improve their understanding of problem-solving process in the context of life sciences. The Intervention was carried out as five learning modules on "Human systems," and each module corresponded to different stages of engineering design practice. Our data analysis investigated the changes of problem-solving process with two independent variables: different genders and robot learning basis. The results showed educational robots can help students more effectively comprehend life science knowledge and understand the problem-solving process. By contrast, there are differences in the problem-solving process between females and males, and the robot learning basis can help students better articulate the problem-solving process. Although this study provided empirical evidence that educational robots can enhance the learning and problem solving skills for primary school children, future studies need to further explore the differences in problem-solving process from multiple perspectives to improve teaching and curriculum design practices.

Keywords: Problems-solving process, Educational robots, Primary school, Life sciences, Engineering design

1. Introduction

In a general sense, problem-solving can be regarded as a kind of mechanical, systematic yet complex skillset to acquire. For instance, solving math problems, which usually have one correct answer, involves a straightforward, logical procedure. However, in daily life, the problems we experience are usually ill-structured and complex, which leads to students' inability to solve problems outside the classroom (Yu et al., 2015). NGSS (2013) proposed that K-12 students should have the opportunity to practice design methods through real experience, and apply scientific knowledge to real world problem-solving. Therefore, more and more research has begun to focus on improving students' ability to solve ill-structured problems through engineering design activities, among which the educational robot is considered to be an effective teaching tool (Jung & Won, 2018; Spolaôr & Benitti, 2017) applied to the design of these projects.

Previous studies have shown that combining engineering design with educational robots can improve students' scientific, mathematical and engineering performance (Bethke Wendell & Rogers, 2013) and develop students' problem-solving skills (Li et al., 2016). However, most research has focused on the combination of simple machines and educational robots and little attention is paid to the changes in students' problem-solving process. Starkweather (1997) argued that cognitive understanding must be included in the curriculum, teaching, and evaluation development of technical and engineering education. It is also more valuable to explore the changes in the problem-solving process of students when dealing with ill-structured problems. The purpose of this study is to develop robot courses based on life science content through the combination of robot education and engineering design. We aim to help students to learn to use engineering design methods to solve practical problems and improve their understanding of the problem-solving process, Based on the above framework of thinking, this study proposes the following research questions:

- Can students learn life science knowledge effectively and improve their understanding of the problem-solving process with robot-based engineering design tasks?
- Are there any differences in problem-solving processes of students in different genders?
- Are there any differences in problem-solving processes of students with different learning basis of robots?

2. Problem-solving process

In the study of Grubbs et al. (2018), design process models for identifying problem-solving strategies were divided into three categories: (1) General engineering design process model (GEDP) (for example, Yu et al. (2015) focused on students' cognitive activities and compared their problem-solving process with the theoretical design process); (2) Professor engineering design process (PEDP), which aims to develop students' design ability to become experts by comparing the cognitive processes of students in different courses (Mentzer et al., 2015; Sung & Kelley, 2019); (3) Cognitive science, such as the study of Wells et al. (2016), directly focuses on the reasoning process of participants. Among the three models, process factor is one of the most critical elements to consider, which is also the key to improving students' problem-solving ability by helping them understand each step involved in a solution and then connect it with real-life situations.

When exploring the problem-solving process of students, however, it was notable that many studies had used the Concurrent think-aloud (CTA) method (Kelley & Sung, 2017; Strimel, 2014; Wells et al., 2016). While this approach can accurately capture the students' short-term thinking process, the extraction of long-term memory status is still a difficulty (Lloyd et al., 1995). Meanwhile, according to Ericsson and Simon (1984), the CTA approach requires students to: (1) Describe design-related thought processes, and (2) Make use of the prior knowledge acquired in the course. As a result, young children are unable to effectively engage in the Think aloud (Think-aloud) approach, which is one of the limitations of the CTA (Van Someren et al., 1994). Some scholars believed that students can achieve short-term and long-term cognition by expressing their ideas in a graphical representation (Ullman et al., 1990), therefore this research adopts the form of chart to allow students to express the problem solving process and analyze it in the sequential pattern.

3. Method

3.1. Participants and procedure

The participants were 69 fifth graders (aged 10-11 years old) recruited from two separate classes of a primary school located in Shanghai, including 36 boys (52.2%) and 33 girls (47.8%). The experiment took place during regular school hours, meaning all participants were taught in two separate classrooms with a similar gender distribution. To ensure the consistency of teaching, the two classes were taught by one of the researchers with identical weekly learning goals as well as curriculum content. Meanwhile, in order to further observe how primary school children implemented their problem-solving strategies, all participants in both classes were randomly assigned to 12 cell groups with two to three students in each group (Figure 1).

Figure 1. Curriculum instruction

Considering the cognitive development of fifth graders, we chose the theme of helping autistic children design robot friends. The course was divided into 5 modules (Table 1). Each module corresponded to different content of life science about human systems and body's sensory. Students were required to participate in didactical activities according to the process of engineering design. The course lasted for 4 consecutive weeks and was 90 minutes in length each week.

Table 1. Course content

Module	Engineering design	Time	Students' challenge
1. I have a robot friend	(1) Identify need or problem	90min	1. Understand the knowledge of autism, help children to establish a healthy psychological concept; 2. Understand the engineering design process; 3. Learning about the composition of human systems and the hierarchy of human structures; 4. Define problems and set goals.
2. Three days to see	(2) Research need or problem (3) Develop possible solution (4) Select the best possible solution	90min	5. Understand the working principle of the ultrasonic sensor; 6. Learning about the body's sensory organs and ways to protect them (eyes, ears, brain); 7. learning to use ultrasonic sensors; 8. Learn the method of sketch design, brainstorm, design the solution, and select the best solution.
3. My emotions	(5) Construct a prototype (6) Test and evaluate the solution (7) Communicate the solution	90min	9. Understand how touch sensors work; 10. Learning how the body feels; Learn simple programming; 11. Build the model according to the solution, test and evaluate; 12. Communication in the group.
4. I need to upgrade	(8) Redesign	45min	13. Improve sketch plan and model design
5. Make more friends	(9) Completion	45min	14. Group communication display, the introduction of works; 15. the completion of robot friends.

3.2. Design

A single group (n = 69) of pre-and post-test quasi-experimental study was adopted in this study. We first compared the participant's overall mastery of subject knowledge through paper-based drawings and written tests before and after the instructional intervention. Later, investigations on each participant's perceived problem-solving process were conducted to analyze the perceived number of steps and paths regarding problem-solving task. To further clarify the effect of intervention on one's problem-solving process, we included "gender" and "robot learning basis" as two independent variables. Robot learning basis referred to whether participants had previous learning experiences with robotics and a questionnaire was designed to identify any differences in robotic knowledge. Aside from quantitative analysis, semi-structured interviews were carried out to supplement the experimental findings.

3.3. Data collection

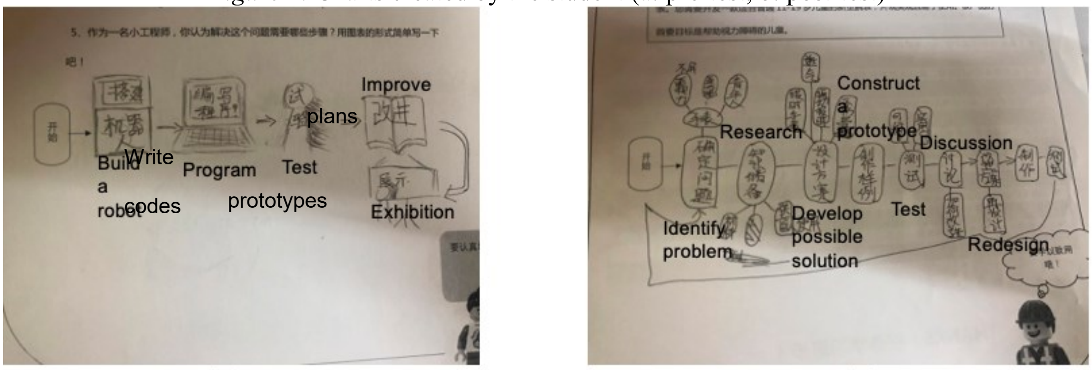
This study mainly collected three forms of data to compare and evaluate the teaching effect and students' cognition of the problem-solving process.

(1) Chart To understand the problem-solving process, participants were required to draw a chart of the problem-solving process as an engineer based on an assigned, specific problem situation before and after the course. In Figure 2, for instance, students outlined their problem-solving process as a straightforward, four-step framework in the pre-test while a more contemplated and sophisticated process is outlined in the post-test.

(2) Subject paper To evaluate the teaching effect of the course, the researchers developed the subject paper, which was composed of background information, life science questions and engineering questions. The background information was to investigate the participant's demographics and robotics learning basis (i.e., whether the student has taken a robotics course prior to the study). The questions were in the form of multiple-choice questions and fill-in-the-blank questions. Life science content accounts for 80% and engineering content accounts for 20%. There are some changes in the sequence of questions and options in the pre-and post-test, but the difficulty and content of the test remained the same. All participants were given ample time to fill in the questions before and after the course.

(3) Interview outline To further analyze the teaching effect of the course, 12 students from the two classes were randomly selected to conduct semi-structured interviews after the course.

Figure 2. Charts created by the student (a: pre-test, b: post-test)



(a) (b)

3.4. Data coding and analysis

The engineering design model proposed by Hynes et al. (2011) is used as the coding scheme in this study to encode the charts drawn by students and identify their problem-solving process. The coding scheme divides the problem solving process into 9 steps, including: (1) Identify need or problem, (2) Research need or problem, (3) Develop possible solution, (4) Select the best possible solution, (5) Construct a prototype, (6) Test and evaluate the solution, (7) Communicate the solution, (8) Redesign, and (9) Completion. We used P1-- P9 to represent the different engineering design steps. To ensure the reliability of coding, charts of 30 students are randomly selected by another researcher for coding (Kappa = .816), showing high consistency. In this study, GSEQ5.1 software is used for sequence analysis of coded data, which is developed by Bakeman and Quera (2015).

Sequence analysis can help researchers effectively identify problem-solving patterns of students (Jung & Won, 2018). In sequence analysis, "Given" represents the current event and "Target" represents the target event, namely the second of two consecutive events. The degree of correlation between the current event and the target event is expressed by adjusted Z-score. Bakeman and Gottman (1997) proposed the formula for adjusted Z-score:

$$\text{Adjusted residual (Z-score)}=\frac{x_{rc}-e_{rc}}{\sqrt{e_{rc}\left(1-\frac{f(c)}{N}\right)\left(1-\frac{f(r)}{N}\right)}} \quad \text{(Bakeman \& Gottman, 1997)}$$

Where X_{re} represents the frequency of the two observed events, e_{rc} represents the expected frequency, f (c) is the total number of events in column c, f (r) is the total number of events in column r, and N is the total number of events. According to the formula, the larger X_{re} is, the larger the Z-score is, indicating a stronger correlation between the two behaviors. For the data collected from the questionnaire, SPSS 21 is used for descriptive statistics in this study.

4. Result

4.1. Academic results and cognition of the problem-solving process

4.1.1. Academic results

Paired *t*-test was adopted in this research to explore students' mastery of subject knowledge before and after the implementation of the course (Table 2). The results showed that there was a significant difference in the subject scores of students ($p < .01$), meaning the post-test scores of students significantly improved as compared with the pre-test scores.

Table 2. Pre-and Post-test of academic results

	N	*M*	*SD*	*t*	Sig. (2-tailed)
Pre-test	69	8.64	1.697	-8.015	.000**
Post-test	69	11.30	2.103		

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

4.1.2. Cognition of problem-solving process

The number of steps in the problem-solving process of students was shown in Figure 3. In the pre-test, students' problem-solving steps mainly focused on 3 to 4 steps while only a few students can write the complete steps. In the post-test, students can generate more steps, mainly focusing on 6 steps. And the number of students who wrote 8 steps also increased. At the same time, this study paired the students' problem-solving steps pre- and post-test with the *t*-test (Table 3), and the results showed that a significant difference in the number of problem-solving steps before and after the implementation of the curriculum ($p < .01$).

Figure 3. Problem-solving steps of students

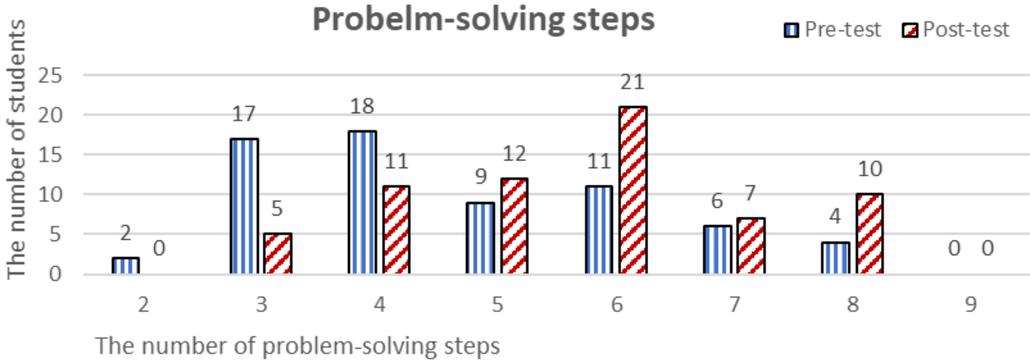


Table 3. Pre-test and Post-test of problem-solving steps

	M	*N*	*SD*	*t*	Sig. (2-tailed)
Pre-test	4.64	67	1.595	-4.097	.000**
Post-test	5.59	66	1.478		

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

To further explore the changes in students' cognition of problem-solving process before and after the curriculum, the initial steps of students' problem-solving process during pre-test and post-test were counted in this study (Table 4). The results indicated that students took a step from P1 (Identify need or problem), P2 (Research need or problem), P3 (Develop possible solution), and P5 (Construct a prototype) as the first step of problem-solving in both pre-test and post-test. In the pre-test, 37% of the students believed that the first step of problem-solving is to "construct a prototype," while only 28% of the students believed that **the first step** of problem-solving is to "Identify need or problem." In the post-test, the proportion of students taking "Identify need or problem" as the starting step was significantly increased (58%), while the proportion of students taking "construct a prototype" as the starting step was significantly reduced (only 1%).

Table 4. The starting step of the problem-solving process

	Pre-test		Post-test	
	N	%	*N*	%
P1	19	28%	38	58%
P2	9	14%	11	17%
P3	14	21%	16	24%
P5	25	37%	1	1%
Total	67	100%	66	100%

To explore the changes of the problem-solving path of most students before and after the curriculum, in this study, the problem-solving process codes of 37% of the students in the pre-test and 58% of the students in the post-test who started with the "Construct a prototype" and "Identify need or problem" were respectively input into GSEQ5.1 to calculate and generate adjusted Z-scores.

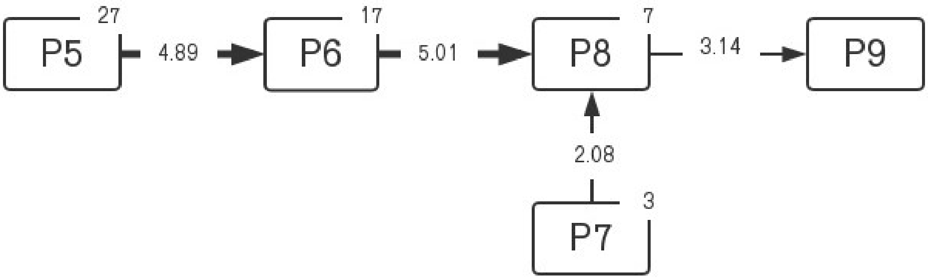
Table 5 is the adjusted Z-score residual table of students' problem-solving process in the pre-test, and Z-score >1.96, meaning there was a significant correlation between their engineering design behaviors. As shown in Table 5, in the pre-test, the steps with significant correlation in students' problem-solving include the following conditions: P5-P6, P6-P8, P7-P8, P8-P9. According to the adjusted Z-score residuals table, most students in this study generated the problem-solving path diagram exhibited by Fig. 4, before the implementation of the course. The nodes in the figure represent different user behaviors. The number in the upper left corner of the node represents the number of people from the current behavior to the target behavior. The connection between behaviors was significant. The arrow represented the direction of the behavior transition. The thickness of the line indicated the degree of correlation of the behavioral connection, while the data on the line was the adjusted Z-score. As shown in Figure 4, in the pre-test, most students chose to start with the "Construct a prototype," focusing on the construction and testing of the model, and make improvements through group discussions to finally complete the design work.

Table 5. Adjusted Z-score residual table for pre-test

	Target								
Given	P1	P2	P3	P4	P5	P6	P7	P8	P9
P1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
P2	0.00	-0.20	0.00	0.00	0.82	0.64	-0.35	-0.76	-0.62
P3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
P4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
P5	0.00	1.48	0.00	0.00	0.00	**4.89****	1.77	-3.50	-3.80
P6	0.00	-0.93	0.00	0.00	-0.97	-4.38	-0.72	**5.01****	1.71
P7	0.00	-0.31	0.00	0.00	-0.26	-1.49	-0.56	**2.08****	0.27
P8	0.00	-0.62	0.00	0.00	1.07	-0.60	-1.09	-2.40	**3.14****
P9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

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

Figure 4. Most students' problem-solving process (Pre-test)



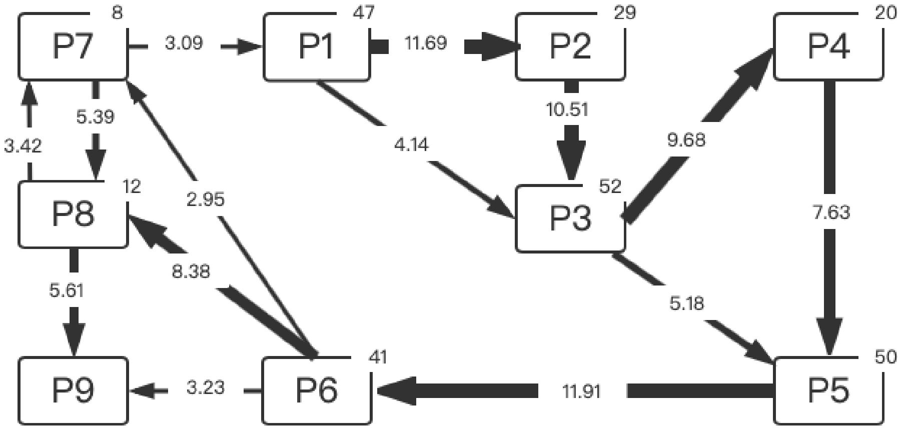
Similarly, Table 6 is the adjusted Z-score residuals table for the problem-solving process of most students in the post-test. As shown in Table 6, in the post-test, the steps with significant correlation in students' problem-solving include the following conditions: P1-P2, P1-P3, P2-P4, P3-P4, P3-P5, P4-P5, P5-p6, P6-P1, P6-P7, P6-P8, P6-P9, P7-P1, P7-P8, P8-P7, P8-P9. According to the adjusted Z-score, most students in this study generated the problem-solving path chart exhibited by Fig. 5, after the course. As shown in Figure 5, in the post-test, the problem-solving process of students started from "Identify need or problem." As shown by the strength of the correlation of students' behaviors, students usually carried out "Research need or problem," "Develop possible solution," "Select the best possible solution," "Construct a prototype," "Test and evaluate the solution" according to the logical order. After the test was completed, most students thought that improvement should be carried out directly according to the test results. Some students also thought that they should discuss and rethink the applicability of problems and requirements, and then continue to iterate between improvement and discussion until the problems are solved.

Table 6. Adjusted Z-score residual table for post-test

Given	Target								
	P1	P2	P3	P4	P5	P6	P7	P8	P9
P1	0.00	11.69**	4.14**	-2.15	-3.52	-3.88	-1.42	-2.69	-1.87
P2	-1.14	0.00	10.51**	-1.70	-2.63	-3.06	-1.12	-2.12	-1.48
P3	-1.68	-2.93	0.00	9.68**	5.18**	-4.50	-0.88	-3.12	-2.17
P4	-0.97	-1.70	-2.50	0.00	7.63**	-1.07	-0.96	-1.81	-1.26
P5	-1.71	-2.99	-4.40	-2.55	0.00	11.91**	-1.69	-2.32	-1.04
P6	3.59**	-2.46	-3.95	-2.49	-3.84	0.00	2.95**	8.38**	3.23**
P7	3.09**	-1.02	-1.49	-0.87	-1.57	-0.74	0.00	5.39**	-0.75
P8	0.05	-0.47	-1.17	-1.55	-0.87	-1.33	3.42**	0.00	5.61**
P9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

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

Figure 5. Most students' problem-solving process (Post-test)



4.2. Problem-solving process of students in different genders

To answer this question, the number of problem-solving steps of boys and girls in pre-and post-test is paired to a *t*-test. As shown in Table 7, there was a significant difference in the number of problem-solving steps for boys before and after the course ($p < .01$), while there is no significant difference found for girls. However, on the whole, the number of problem-solving steps before and after the implementation of the curriculum was higher for girls than that for boys. In the pre-test, there is a huge difference in the number of problem-solving steps between girls and boys. After the course, however, the number difference of problem-solving steps between the two genders was relatively smaller in the post-test.

Figure 6. Problem-solving process of students in different genders (a: female, b: male)

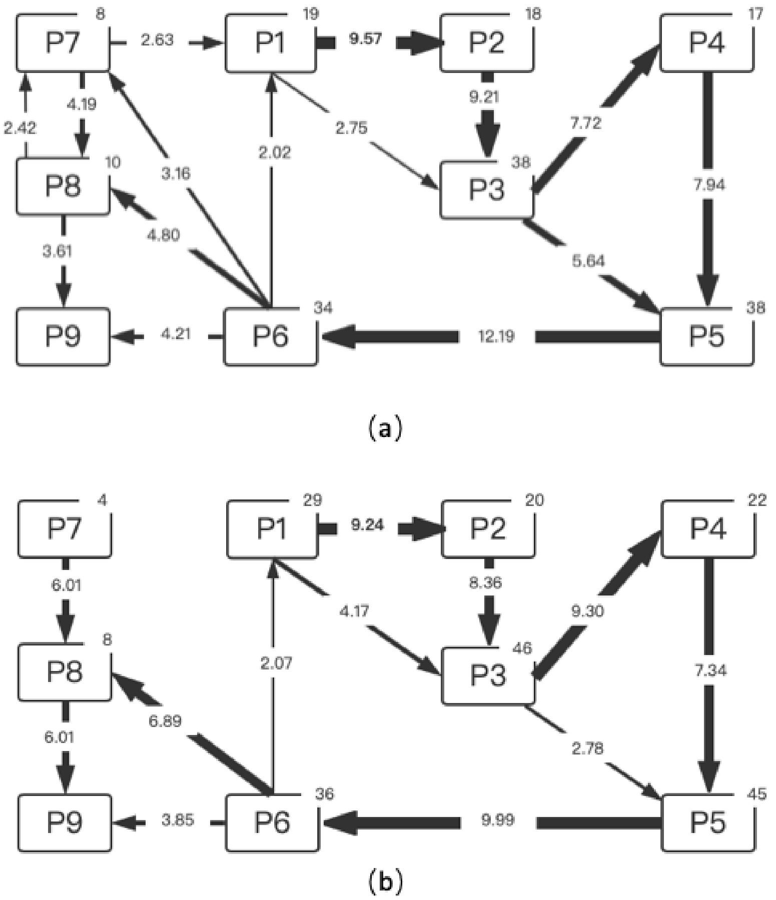


Table 7. Problem-solving steps of students in different genders

		M	*N*	*SD*	*t*	Sig. (2-tailed)
Male	Pre-test	4.17	35	1.272	-4.460	.000**
	Post-test	5.51	35	1.442		
Female	Pre-test	5.16	31	1.772	-1.476	.150
	Post-test	5.68	31	1.536		

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

In this study, the number of problem-solving steps of boys and girls was carried out with an independent sample *t*-test to explore whether gender is the key factor affecting problem-solving. The results of Levene's variance test showed that (Sig = .719), the variance was homogeneous, and there was no significant difference in the number of problem-solving steps between the two genders (Sig = .658), indicating that gender was not statistically significant in affecting the number of steps they took in problem-solving process. Additionally, we have compared the problem-solving process of boys and girls in this study. As shown in Figure 6 (a) and (b), boys and girls show a strong correlation between the first 6 steps of problem-solving (P1-P2, P2-P3, P4-P5, P5-P6) after taking part in the course. The difference was that compared with boys, girls still have a strong correlation between P3 (Develop possible solution) and P5 (Construct a prototype). Some girls did not distinguish very well between P4 (Select the best possible solution) and P3. For the last three stages of problem-solving (P7, P8, P9), boys and girls produced different results. Girls were more aware of the importance of P7 (Communicate the solution) and thought about the applicability of solutions than boys. P7-P1, P8-P7, and P6-P7 were correlated to some extent.

4.3. Problem-solving steps of students with different robot learning basis

This study takes whether students have the robot learning basis as an independent variable to explore the influences of different experiences on the understanding of the problem-solving process. In this study, the number of problem-solving steps of students who have the robotics learning basis and students who do not have the robotics learning basis are paired with a *t*-test. As shown in Table 8, there was a significant difference in the number of problem-solving steps for students who have learned robotics courses before and after the implementation of the course ($p < .01$), while there was no significant difference in the number of problem-solving steps for students who don't have robotics learning basis. On the whole, regardless of previous learning experiences, the number of problem-solving steps among students was very close in the pre-test. After the intervention, however, the number of problem-solving steps for the students with robotics learning basis was higher than that for the students without robotics learning basis.

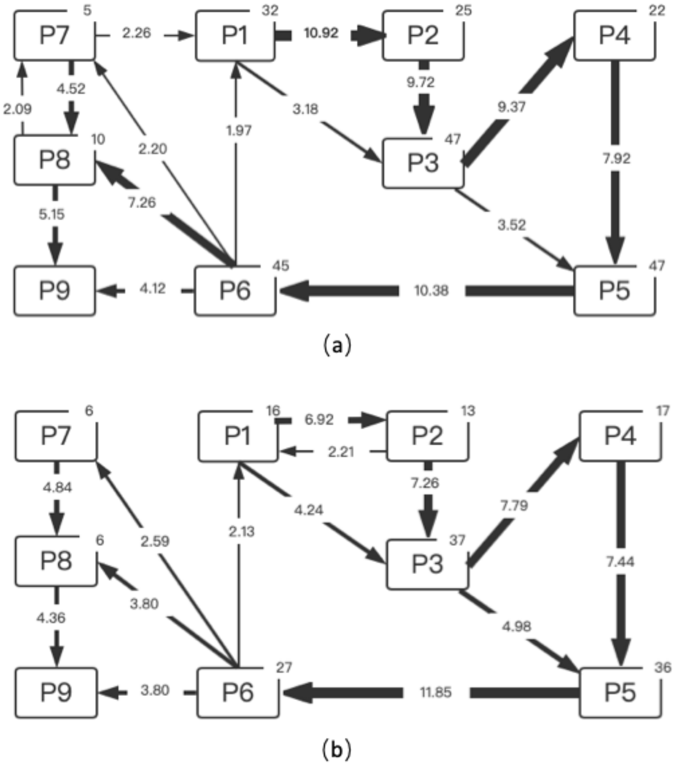
Table 8. Problem-solving steps of students in different robot learning basis

		M	*N*	*SD*	*t*	Sig. (2-tailed)
Without robot learning basis	Pre-test	4.68	28	1.786	-1.565	.129
	Post-test	5.36	28	1.660		
With robotics learning basis	Pre-test	4.66	38	1.419	-3.698	.001**
	Post-test	5.76	38	1.324		

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

An independent sample *t*-test was also conducted on the number of problem-solving steps between students with and without robotics learning basis. The results of the Levene's variance test (Sig = .102) showed homogeneity of variance. There was no significant difference in the number of problem-solving steps between different robotics learning basis (Sig = .273). Therefore, whether students have robot learning experiences does not statistically significantly in affect the number of steps they took in problem-solving process. When analyzing the actual problem-solving process, as shown in Figure 7 (a) and (b), students under both experience conditions have a strong correlation between the first six steps of problem-solving (P1-P2, P2-P3, P3-P4, P4-P5, P5-P6). Moreover, students who had learned robotics courses have a clearer logical relationship with these steps, and the correlation between these steps was higher than those without robotics learning basis. As for the last three stages of problem-solving (P7, P8, and P9), students who had a robotics learning basis pay more attention to the stage of P7 and P1, and there is a certain correlation between P7-P1 and P8-P7.

Figure 7. Problem-solving process of students with different robot learning basis (a: with robot learning basis, b: with-out robot learning basis)



4.4. Interviews

This study conducted semi-structured interviews with students from two aspects of problem-solving and course content. The results were as follows:

4.4.1. Course content

It can be inferred that participants generally think it is difficult to design and build solutions by themselves. Besides this, students think using robots to learn life science knowledge is very interesting. Most of the students said they learned a lot about building and programming robots, as well as about human systems and organs, and some said they learned how to carry out a project to design and solve problems.

Researcher： Do you think the course content is difficult? If so, which parts do you find difficult?

Student： A little bit, the design and construction of our own hands, no drawings, a little hectic. (Student 1). One thing is, it's too hard for me to design my solution and make it workable. (Student 6)

Researcher: Do you think the course content is interesting? What have you learned?
Student: Interesting, I learned how to carry out a project, design, production. (Student 3) Interesting, I know the steps to solve the problem, human organs and robot organs. (Student 1)

4.4.2. Problem-solving process

In terms of problem-solving, most students have a clear process in-mind and can describe the complete steps in their solution plans. Students think that the way to solve problems learned in the course is applicable to other situations in everyday life.

Researcher： Through this course， what steps do you think should be taken to solve the problem?

Student：Identify problems, build knowledge, develop possible solutions, build models, discuss, redesign, and work until you find the best solution. (Student 1). Identify questions, think about what materials you have, interview, produce, test. (Student 4)

Researcher: Do you think the knowledge gained in this training course helps you to solve problems in your everyday life?

Student: There is a certain help when solving math problems, I used to do it directly, now I know that I can first look at the problem type, and then think about which solution to implement, and then choose the best one (Student 6). Yes. Now if I have some problems, I know where to start with. (Student 4)

5. Discussion

5.1. Academic results and cognition of the problem-solving process

5.1.1. Academic results

Statistics of this study show that educational robots can help students learn life science knowledge, and students' test scores have been significantly improved in the post-test. From the interview results, we can see that students are very interested in attending robot design courses, which also improves their motivation for learning courses and classroom participation. Also, this teaching adopts student-centered inquiry and "learning by doing" to help students better understand life sciences content. This is consistent with previous research findings, which show that robotics courses can encourage students to think and discuss, and at the same time, improve students' learning motivation and promote classroom learning (Cukurbasi & Kiyici, 2018) to improve learning efficiency (Bethke Wendell & Rogers, 2013). However, the research found that some students still fail to understand the content of this part in the post-test, resulting in low scores. Additionally, the interview results show that some students might have difficulty memorizing the details of course content, which is the main reason why they failed to score effectively. In terms of course design, some course content is relatively abstract and not closely related to robot tools. Moreover, course content itself is difficult, which leads to students' inability to master course contents.

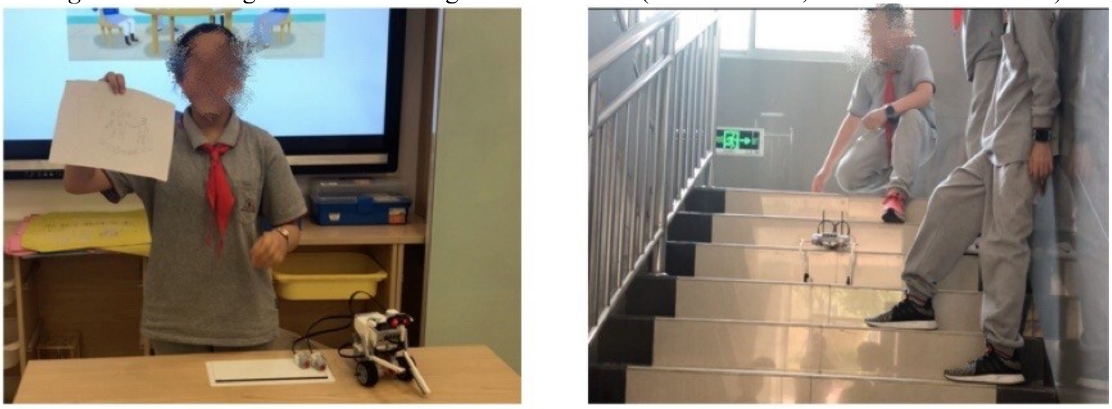
5.1.2. Cognition of problem-solving process

It is evident that students can generate more steps for problem-solving after the course. From the results of specific charts, in the post-test, students' steps of problem-solving become more systematic and logical, and the process of work design is more in line with the process of engineering design. Previous studies have shown that robot education can improve students' problem-solving ability (Ilori & Watchorn, 2016; Li et al., 2016). From the perspective of the change of the problem-solving process, this study proves the improvement of students' problem-solving ability.

Sung and Kelley (2019) believed that sequence analysis can successfully describe the problem-solving process of young learners. According to the results of sequence analysis, in the pre-test, most students start with the construction of the prototype and focus on the design and improvement of the model. However, little attention has been paid to other steps of problem-solving. According to the research of Mentzer et al. (2015), compared with experts, novice problem solvers usually spend less time on problem definition and developing possible solutions. Meanwhile, it is also found that students tend to ignore the test (P6) stage in the design process (Kelley et al., 2015), which is consistent with the results of this study. But after teaching, students' problem-solving process is improved. The process is more logical and systematic in the post-test, and students pay attention to the importance of steps such as P1 (Identify need or problem), P2 (Research the problem) and P3 (Develop possible solution), they begin to design iteratively around problems and requirements. The results of the interview show that most students can understand the process of problem-solving, describe the steps of problem-solving, and use this process as a framework of thinking to solve similar problems encountered.

In this study, robot education is carried out to help students understand the process of problem-solving. However, this study found that in the post-test, there are still some students' problem-solving methods centered on "Develop possible solution" (24%) and "Construct a prototype" (1%). This study compared these students (25%) with the most students (58%). In Figure 8 (a), students designed a "robot friend" equipped with sensors that could interact with autistic children. And in Figure 8 (b) students designed a robot that can descend stairs. The design of the second group is more creative because they are constantly modifying and building the robot. However, they cannot solve problems well, which is also the reason why they ignore "identifying problems and requirements." On the contrary, the design of first group can solve problems better.

Figure 8. The design of different categories of students (a: robot friend, b: descend stairs robot)



(a) (b)

5.2. Differences in problem-solving process among students of different genders

Both boys and girls showed positive results in the problem-solving process after learning the course. Through the comparison of the results of the sequence analysis of the problem-solving process between male and female students, this study finds that there are still some differences in the way boys and girls approach problems. Strimel (2014) found in his study that girls spend more time communicating and designing solutions than boys, and believed that effective communication can enhance the ability to solve problems. That is why girls can write more steps of the problem-solving process than boys in the pre-test. The focus is on the two steps of "Develop possible solution" and "Communicate the solution." This is consistent with the results of this study. In the post-test, girls prefer to improve the design of work through continuous communication and discussion in the process of redesign and test evaluation, while boys pay less attention to the importance and iteration of communication and discussion in test evaluation and redesign. Besides, this study finds that compared with boys, girls pay more attention to the importance of "Identify need or problem" and think about whether the solution or the design of the work meets the needs of the problem. However, this study suggests that too much emphasis on design solutions may lead to girls' unclear distinction between "Develop possible solution" (P3) and "Select the best possible solution" (P4).

5.3. Differences in problem-solving processes of students with different learning basis of robots

The research results show that students of both learning basis categories can achieve positive changes. However, students with the basis of robot learning have made greater progress after intervention, which is manifested in that they can write more problem-solving steps, and their problem-solving process is more logical and systematic. This study found that the basis of robot learning can help students better understand the process of problem-solving and express the steps of problem-solving. This study tries to find valid evidence to support the following viewpoints from relevant studies in terms of knowledge and skills. (1) Students without the basis of robot learning lack prior knowledge related to robots. When students are required to think about how to carry out projects, they are often unable to describe the process (Barak & Zadok, 2009). On the contrary, the more conceptual knowledge students have, the better their project performance will become to effectively solve problems (Fan et al., 2018). (2) Teamwork, communication, and problem-solving are the most common skills of students trained by robots (Spolaôr & Benitti, 2017). The learning experience basis of robots helps students have a good knowledge and skill base, and therefore, helps students better comprehend the problem-solving process.

The results of sequence analysis show that students with the basis of robot learning have a clearer problem-solving process than students without the basis of robot learning, strong correlations are formed between P1-P2, P2-P3, P3-P4, and P4-P5, and attention is paid to the important role of "Communicate the solution" in the process of "Redesign," and the applicability of works or solutions to problems and requirements is considered. Strimel (2014) believed that the problem-solving steps of non-sequent participants were often chaotic and did not plan to solve the problem before constructing the prototype. By comparing the problem-solving paths of students of different categories, this study found that students of the two categories showed different degrees of "non-sequent participants." However, after the course training, the problem-solving processes of students in both categories begin to shift to the direction of "sequential participants" that they can solve problems in a logical order.

6. Conclusion

Based on the life science content and engineering design practice, this study designed a set of robot courses and explored their influence on students' problem-solving process. The study found that "when life science meets robot education," the presence of a tangible robot in teaching can help students learn life science knowledge and understand the process of problem-solving. However, there are some differences in the problem-solving process of different categories of students. For instance, male students tend to ignore communication and thinking about the applicability of solutions, while female students can hardly distinguish between designing solutions and choosing the best one. Ultimately, the foundation of robot learning can help students better understand the process of problem-solving and generate more systematic and logical problem-solving methods. Therefore, how to pay attention to the differences between students of different genders and help students with weak robot learning foundation to improve their understanding of the problem-solving process is the focus of future research, and also the part that should be improved in the robotics course of this study.

It is useful to teach students the basic knowledge of relevant scientific concepts and related concepts of problem-solving (Barak & Zadok, 2009), but it should be done in a flexible method rather than strict teaching. For instance, we found robotics offers children decent opportunities to apply mechanical design and engineering knowledge to further developing their skills in problem-solving. In other words, the learning and reflection of problem-solving practices could be effectively carried out through hands-on design experiences with robots. Future research might expand the scope of curriculum content to other subjects to engage children in developing problem-solving skills through designing and building up a tangible robot.

There are still some limitations in the implementation of this study. Firstly, the duration time of the course might be insufficient, which leads to an insignificant effect on the change of problem-solving process. Secondly, this study does not investigate the results of students' problem-solving from multiple perspectives, such as work analysis, sketch design and thinking, etc. Finally, in this study, the researcher is both teacher and interviewer. The researchers' multiple identities inevitably lead to subjective bias on the results of student interviews and questionnaires that researchers will unconsciously guide students to answer or fill in the blanks in a positive direction. Future research should involve different researchers in teaching and student interviews.

Reference

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An Introduction to sequential analysis* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511527685

Bakeman, R., & Quera, V. (2015). *Generalized sequential* [Computer software]. https://www.mangold-international.com/en/products/software/gseq

Barak, M., & Zadok, Y. (2009). Robotics projects and learning concepts in science, technology and problem solving. *International Journal of Technology and Design Education, 19*(3), 289-307. https://doi.org/10.1007/s10798-007-9043-3

Bethke Wendell, K., & Rogers, C. (2013). Engineering design-based science, science content performance, and science attitudes in elementary school. *Journal of Engineering Education, 102*(4), 513-540. https://doi.org/10.1002/jee.20026

Cukurbasi, B., & Kiyici, M. (2018). High school students' views on the PBL activities supported via flipped classroom and LEGO Practices. *Educational Technology & Society*, *21*(2), 46–61.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. The MIT Press.

Fan, S.-C., Yu, K.-C., & Lou, S.-J. (2018). Why do students present different design objectives in engineering design projects? *International Journal of Technology and Design Education, 28*(4), 1039-1060. https://doi.org/10.1007/s10798-017-9420-5

Grubbs, M. E., Strimel, G. J., & Kim, E. (2018). Examining design cognition coding schemes for P-12 engineering/technology education. *International Journal of Technology and Design Education, 28*(4), 899-920. https://doi.org/10.1007/s10798-017-9427-y

Hynes, M., Portsmore, M., Dare, E., Milto, E., Rogers, C., Hammer, D., & Carberry, A. (2011). *Infusing engineering design into high school STEM courses*. https://digitalcommons.usu.edu/ncete_publications/165

Ilori, O., & Watchorn, A. (2016). Inspiring next generation of engineers through service-learning robotics outreach and mentorship programme. *International Journal of Advanced Robotic Systems, 13*(5), 1729881416663372. https://doi.org/10.1177/1729881416663372

Jung, S., & Won, E.-S. (2018). Systematic review of research trends in robotics education for young children. *Sustainability, 10*(4), 905. https://doi.org/10.3390/su10040905

Kelley, T., & Sung, E. (2017). Examining elementary school students' transfer of learning through engineering design using think-aloud protocol analysis. *Journal of Technology Education, 28*(2), 83-108.

Kelley, T. R., Capobianco, B. M., & Kaluf, K. J. (2015). Concurrent think-aloud protocols to assess elementary design students. *International Journal of Technology and Design Education, 25*(4), 521-540. https://doi.org/10.1007/s10798-014-9291-y

Li, Y., Huang, Z., Jiang, M., & Chang, T.-W. (2016). The Effect on pupils' science performance and problem-solving ability through Lego: An Engineering design-based modeling approach. *Educational Technology & Society, 19*(3), 143-156.

Lloyd, P., Lawson, B., & Scott, P. (1995). Can concurrent verbalization reveal design cognition? *Design Studies, 16*(2), 237-259. https://doi.org/10.1016/0142-694X(94)00011-2.

Mentzer, N., Becker, K., & Sutton, M. (2015). Engineering design thinking: High school students' performance and knowledge. *Journal of Engineering Education, 104*(4), 417-432. https://doi.org/10.1002/jee.20105

Next Generation Science Standards (NGSS) Lead States. (2013). *Next generation science standards: For states, by states*. The National Academies Press.

Spolaôr, N., & Benitti, F. B. V. (2017). Robotics applications grounded in learning theories on tertiary education: A Systematic review. *Computers & Education, 112*, 97-107. https://doi.org/10.1016/j.compedu.2017.05.001

Starkweather, K. N. (1997). Are we thinking to achieve? *Technology Teacher, 57*(2), 5-6.

Strimel, G. J. (2014). *Engineering design: A Cognitive process approach.* Old Dominion University. https://doi.org/10.2577/zzbj-b616 https://digitalcommons.odu.edu/stemps_etds/99

Sung, E., & Kelley, T. R. (2019). Identifying design process patterns: A Sequential analysis study of design thinking. *International Journal of Technology and Design Education, 29*(2), 283-302. https://doi.org/10.1007/s10798-018-9448-1

Ullman, D. G., Wood, S., & Craig, D. (1990). The Importance of drawing in the mechanical design process. *Computers & Graphics, 14*(2), 263-274. https://doi.org/10.1016/0097-8493(90)90037-X

Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. (1994). The Think aloud method: A Practical approach to modelling cognitive. *AcademicPress*, *11*. http://hdl.handle.net/11245/1.103289

Wells, J., Lammi, M., Gero, J., Grubbs, M. E., Paretti, M., & Williams, C. (2016). Characterizing design cognition of high school students: Initial analyses comparing those with and without pre-engineering experiences. *Journal of Technology Education, 27*(2), 78-91. http://doi.org/10.21061/jte.v27i2.a.5

Yu, K.-C., Fan, S.-C., & Lin, K.-Y. (2015). Enhancing students'problem-solving skills through context-based learning. *International Journal of Science and Mathematics Education, 13*(6), 1377-1401. https://doi.org/10.1007/s10763-014-9567-4

Whose Spatial Ability Benefits from Learning with 3D Design? From the Perspective of Learning Analysis

Shouchao Guo[1], Xiao Wang[1], Wenbo Deng[2], Jialing Hong[1], Jiawen Wang[1] and Yonghe Wu[1*]

[1]Department of Education Information Technology, East China Normal University, China // [2]School of Education, Shaanxi Normal University, China // shouchaoguo@126.com // 3245794069@qq.com // wenbolz@163.com // 51204108009@stu.ecnu.edu.cn // wangjiawen_123@163.com // yhwu@deit.ecnu.edu.cn

[*]Corresponding author

ABSTRACT: Three-dimensional (3D) design can improve students' spatial ability, but the research on the differences of spatial ability development after 3D design training for students with different initial spatial ability is not unified. The ability-as-enhancer hypothesis and the ability-as-compensator hypothesis explain the performance differences of students with different initial spatial abilities in different situations. However, the existing research has not formed a consistent conclusion, which makes students lack of fine guidance, and it is difficult to achieve good spatial ability training effect. This study first explored the differences of students' performance under different educational interventions, and verified the value of process data in the cultivation of spatial ability. Then, we collected more students' data, discussed the improvement of students' spatial ability by 3D design with different initial spatial ability, and tried to explain the difference of students' performance by students' 3D design behavior. We found that different educational interventions can affect students' task participation, and then the effect of spatial ability training. Students with different initial spatial abilities still have significant differences in spatial ability after 3D design, but there is no significant difference in the improvement of spatial ability, and no difference in the data of 3D design operation process. Through cluster analysis, this study also found five types of students in the process of 3D design. There are significant differences in the pre-test, post-test only among some types of students. This study provides a reference for the training effect evaluation of students with different initial spatial abilities.

Keywords: 3D design, Spatial ability, Learning analysis, Ability difference

1. Introduction

Human beings live in the space environment, and their survival and development are carried out through the exchange of material and energy with the space environment. In contact with the environment, people must have the ability to judge spatial orientation and determine the spatial relationship and structure of geographical things. As one of human's basic intelligence, spatial ability plays an important role in human survival and development. Many studies have shown that spatial ability are highly correlated with the performance of science, technology, engineering, and mathematics (STEM) subjects (Lubinski, 2010; Sorby et al., 2013). Although spatial ability is one of the most researched factors of human cognitive function (Carroll, 1993), the concept of spatial ability has not been unified yet, and the measurement and testing methods of spatial ability cannot fully measure spatial ability (Höffler, 2010). Spatial ability has always been of secondary interest in the research of human intelligence.

Many studies have found that no matter how students' previous skills, experience, grades, etc. are, their spatial ability can be improved after training (Šafhalter et al., 2020). Ability-as-enhancer hypothesis and ability-as-compensator hypothesis are often used to explain differences in students' spatial ability improvement. Mayer and Sims (1994) believe that students with high spatial ability should benefit from animation in particular because they have sufficient cognitive ability to construct a mental model. Hays (1996) believe that students with low spatial ability should benefit from explicit graphical representation because it is difficult for them to construct their own visualization psychologically. Due to the lack of a perfect theoretical framework and the diversity of teaching design, there is a certain contingency in the research, so that it is impossible to form an effective and generalizable teaching practice strategy. Existing research is still more about exploring the universal laws of education through collective teaching, but students' cognitive characteristics or differences have not received enough attention. The students' spatial ability tendency, learning preference, motivation and environment are interfering with each other (Hays, 1996). According to the Aptitude-by-treatment interaction theory (ATI), there are complex interactions between students and teaching strategies. While some teaching methods are generally effective, they may not be effective for students with other characteristics (Mcleod et al., 1977). Students with different spatial abilities have different cognitive loads, prior knowledge, learning styles, learning interests, etc. They also have different behavioral characteristics when solving spatial tasks. After students with different spatial abilities use 3D design, can their spatial abilities be improved? Are there any differences in their improvement values? Is there any difference in 3D design behavior for students with different initial abilities? Are there any differences in the improvement of spatial ability among students of different learning types?

2. Literature review

2.1. Spatial ability and its development

Linn and Petersen defined spatial ability as a skill involving representation, transformation, generation, and extraction of symbols and non-verbal information, and they proposed three factors of spatial ability: spatial perception, mental rotation, and spatial imagination (Linn & Petersen, 1985). However, there is no consensus on whether the improvement of students' spatial ability is durable and transferable through training (Heckman & Masterov, 2007; Sims & Mayer, 2002).

At present, the spatial ability of students is mainly measured by methods such as mental rotation test, origami test, mosaic pattern test, etc. The reliability and effectiveness of measurement need to be improved. For one thing, there is no unified definition of spatial ability, and there is no comprehensive measurement scale for each element of spatial ability and comprehensive experience in use. For another, the difference between spatial ability test tools and test environment will also affect students' performance. For example, computer-based spatial ability test can eliminate the male advantage in mental rotation test (Monahan et al., 2008). Regarding whether there are gender differences in the cancellation of the time limit of the mental rotation test, different researchers have found different results (Masters, 1998; Vandenberg & Kuse, 1978). However, when the test time is further shortened, women will use the method of guessing questions to complete the test task, and the gender advantage of men will be more significant (Voyer et al., 2004). Last but not least, Larson pointed out the difference between the dynamic real world and the two-dimensional static test. Paper-pencil test and standardized test restrict the exploration of spatial ability (Larson, 1996). Moreover, different spatial problems have different solutions, and so does the teaching environment. In turn, the differences in cognitive processes and task-solving strategies (Chien & Chu, 2018) also pose challenges to the consistency research of spatial ability diagnosis.

Gender differences, age differences, and strategy differences in spatial ability have been extensively studied, but the reasons for the differences have not been fully explained. Teaching interventions based on differences in spatial ability cannot be widely promoted. The spatial ability test only provides the test results, ignoring the students' efforts and the improvement of logical thinking ability in the task-solving process, and cannot show the problems of students' learning input, learning strategies, knowledge and skills application, etc. Moreover, the teaching intervention focuses on the evaluation and feedback of the overall performance of the students, ignoring the diagnosis and personalized feedback of the individual behavior of each student.

2.2. Spatial ability difference and its intervention

In the early studies of spatial ability, psychologists and educational researchers found gender differences in spatial ability (Baenninger & Newcombe, 1989; Voyer et al., 1995). However, gender differences in spatial ability are also heterogeneous (Linn & Petersen, 1985; Voyer et al., 1995). Men have a dominant advantage in mental rotation (Deno, 1995; Linn & Petersen, 1985; Voyer et al., 1995), while women are more dominant in spatial positioning and perception speed.

In recent years, the role of students characteristics or individual differences in spatial ability training has attracted more and more attention (Höffler & Leutner, 2011; Meijer & Broek, 2010). The ability-as-enhancer hypothesis believes that students with high spatial ability can use less time to extract spatial information and can gain greater gains from 3D design (Huk, 2010). The ability-as-compensator hypothesis believes that 3D design can help students with low spatial ability build a 3D model, without affecting students with high spatial ability or increasing their irrelevant load (Höffler & Leutner, 2011; Hays, 1996; Huang & Lin, 2017; Lee & Wong, 2014). However, the improvement of different spatial ability elements of students is also different. At present, researchers have a consensus that stereotypes in spatial ability will affect students' spatial task performance (Ortner & Sieverding, 2008; Sharps et al., 2010). For example, when there are gender differences in the process of guided mental rotation tasks, the gender differences in performance on mental rotation tasks will increase (Ortner & Sieverding, 2008).

When students receive correct, immediate and personalized feedback in the process of spatial problem solving, they tend to be more motivated to participate in learning activities (Kleij et al., 2012). Compared with paper and pencil tests, procedural data can provide feedback on the thinking process for students and teachers (Whitelock, 2009). Automated data tracking, collection, and storage can avoid the Hawthorne effect of students and the expected effect of teachers. It also avoids standardized tests that rely on language and mathematical logic skills, as well as the restrictions caused by controlled experiments, data reasoning and induction, making it easier to convert the evaluation results into implementable, generalizable, and replicable teaching suggestions.

2.3. 3D design and spatial ability

The 3D virtual environment has unique advantages in simulating the authenticity, interactivity, and visibility of the objective world. 3D design through in-depth integration with traditional education, builds a personalized, interesting, and open comprehensive innovative practical teaching mode. 3D design can make abstract knowledge concrete and help students master science, technology, engineering, mathematics and other knowledge. Different from the methods of training spatial skills such as engineering drawing and sketch training, 3D design provides students with clearer object visualization (Blikstein et al., 2017). Not only can it help students spend less time creating models and improve the accuracy and completeness of the models (Snyder et al., 2014), but it can also help students learn how to solve problems (Blikstein et al., 2017) and cultivate students' creativity (Eisenberg, 2013). In the 3D design process, the choice of graphics, the splicing of graphics, the combination of graphics, and the rotation of graphics all reflect the students' spatial ability. Many scholars have proved that 3D modeling can improve students' spatial ability (Gerson et al., 2001; Koesa & Karakus, 2018).

3D design is not only a tool and means for cultivating spatial ability, but also an important scenario for spatial ability evaluation. In addition to traditional spatial ability evaluation methods such as observation, interviews, questionnaire surveys, and self-evaluation, methods such as work design schemes, operation logs, and screen records have also been applied (Wu et al., 2018). Operating habits such as the number of 3D design operations and operating time are also commonly used to evaluate the teaching effects of 3D design (Al-Ahmari et al., 2018; Barber et al., 2016). 3D design can not only be used for spatial ability training, 3D design process data is also the explicit data of students' spatial thinking, which can be used to evaluate students' spatial ability (Wu et al., 2020). Therefore, we plan to use 3D design tools to cultivate students' spatial ability, and diagnose the improvement of different types of students' spatial ability based on process data, and provide support for teaching decision-making and students' personalized intervention.

2.4. Research hypothesis

The basic assumption of this study is that in the process of 3D design, students' operation behavior can be divided into many types, and different types of operation behavior will get different spatial ability training effect. Students with high initial spatial ability and students with low initial spatial ability will produce different operation behavior data in 3D design, so as to achieve different spatial ability promotion. Specifically, we will ask the following questions:

Research question 1: Compared with paper materials, will 3D printed models affect the training effect of 3D design on spatial ability? Will different educational interventions affect students' behavior?

Research question 2: What is the difference between the 3D design behavior of students with high spatial ability and students with low spatial ability? Is the improvement of space capacity related to operational behavior?

Research question 3: In 3D design, according to the 3D design operation behavior, what types can we divide students into, and what are the improvement differences, in spatial ability, for these students?

3. Methods

3.1. Participants

We conducted two experiments on 22nd May 2020 and 20th October 2020. Prior to the study, we obtained ethical review approval, and participants obtained informed consent and voluntarily signed the consent form.

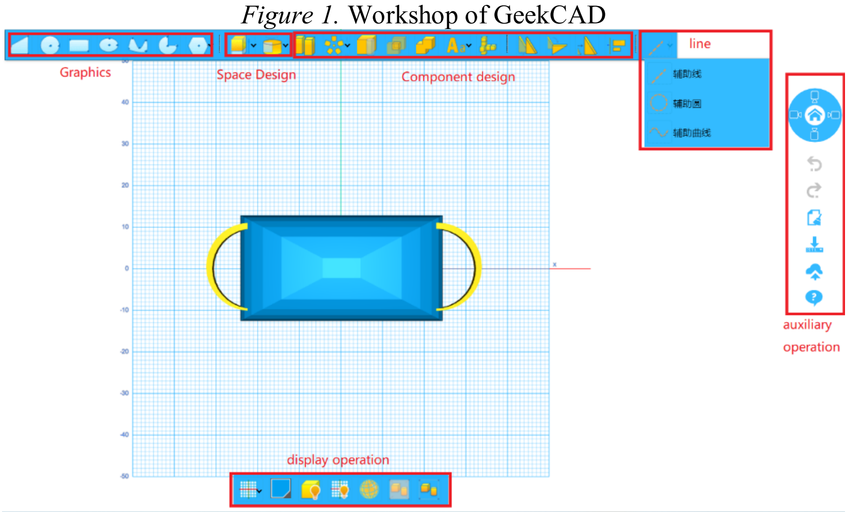
This study does not require prior knowledge or computer skills of the participants. In the first round of the experiment, we selected two classes in grade one of a middle school in Lanzhou City, Gansu Province, China. A total of 97 students participated in the experiment, including 51 students in the experimental group and 46 students in the control group. In the second round of experiment, we conducted this research in two classes of the first grade in a middle school in Lanzhou, China. A total of 88 students participated in this experiment. The students who do not participate in this experiment will use separately assigned accounts to participate in the course normally, but the operation process of the two students will not be recorded and analyzed.

This course was taught by an experienced male information technology teacher. All students completed the research tasks. However, it needs to be explained that because the object of this study cannot represent the level of high school students, the purpose of this experiment is to explain the differences in the operation of different types of students, and the research results need to be further verified to extend to all high school students.

3.2. Materials

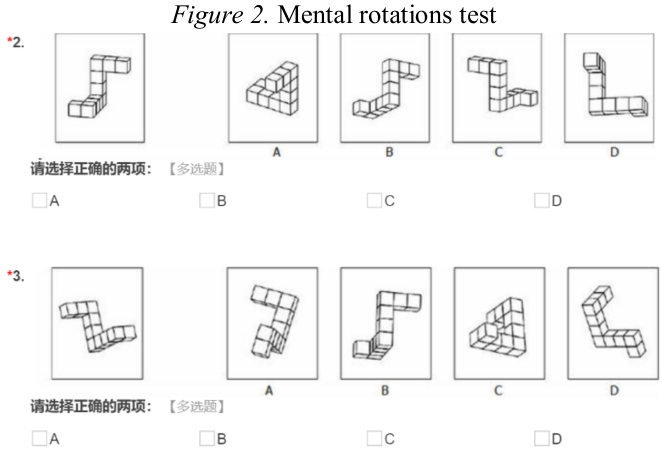
Considering the ease of use and difficulty of 3D design software, geekCAD, a browser-based 3D design tool, was selected for this research. As shown in Figure 1, geekCAD includes seven areas, the commands related to 3D object operation are at the top, and the other areas are system auxiliary functions. When students are designing in 3D, all their operations on the platform will be recorded.

Figure 1. Workshop of GeekCAD



We choose the mental rotation test score as the student's spatial ability, and use the mental rotations test (Vandenberg & Kuse, 1978) shown in Figure 2 for testing. Each question includes a total of five graphics, of which the first graphic is the original graphic. Students need to determine which two of the following four graphics (labeled with ABCD) can be obtained by rotating the original graphic. There are two correct options for each question, and only the students who choose two correct answers will get one point. Checking one answer or selecting zero answers is counted as zero points. Each student has six minutes to complete the test.

Figure 2. Mental rotations test



3.3. Research procedure

Referring to Boucheix and Schneider (2009), we designed two rounds of experiments to explore the differences of students' operation behavior under different educational intervention conditions and the behavior differences of different types of students in 3D design.

The process and duration of the two experiments are consistent with each other. All participants completed the task in four steps. First (10 minutes), the teacher described the test process and organized the students to participate in the mental rotation test. Then (15 minutes), the teacher explained the function and operation method of geekcad. The students tried to make a water cup and get familiar with the operation method of geekcad. Then (80 minutes), the teacher asked the students to design Zhongshan Bridge. Zhongshan Bridge is a famous scenic spot of the Yellow River in China. After confirming that the students have no problem with the task, ask them to complete the design task independently. Students can consult the materials provided by teachers at any time during the design process. Finally (15 minutes), the teacher commented on the students' works and invited them to participate in the mental rotation test again. The first round of experiment was divided into experimental group and control group. The knowledge of the experimental group and the control group are the same, including the introduction of Zhongshan Bridge and three view drawing. 88 students participated in the second round of experiment. The only difference is that in the first round of experiment, we provided the 3D model of Zhongshan Bridge for the experimental group and the paper three view data for the control group. In the second round, we provided all the students with 3D models.

The first round of experiments proved the value of operational behavior in spatial ability evaluation. However, due to the small number of participants in the first round of the experiment, the classification of students may lack credibility. Therefore, in order to further explore the differences in learning performance of different types of students, we carried out a second round of experiments under the same educational intervention. The task of different difficulty and different situation will affect students' learning state. We only measure students' spatial ability, but not their academic performance (Hegarty & Sims, 1994). However, it should be noted that the order of questions before and after the mental rotation test is confused, which avoids the students' practice effect.

3.4. Data collection

In this study, xAPI specification is used to automatically collect students' click stream data during 3D design on geekcad. Each click of students will generate a series of relevant data, including operators, coordinate points, operation objects, operation commands, results, etc. For example, when coloring a model, xAPI can automatically collect data such as which student added which color to which object at what time, and record the mouse or keyboard input data in the process. For xAPI data collection mechanism and students' 3D design behavior specification, please refer to research results of Wu et al. (2020). Response latency, response frequency, and invested time are performance factors often considered in dynamic spatial ability test (Contreras et al., 2007). Furthermore, referring to the research on learning behavior engagement in online learning (Fredricks et al., 2004; Kim et al., 2016), we identified three types of students' 3D design behavior, including nine indicators, such as the number of operations, the type of operations, the duration of investment, and the maximum time interval. The explanation of each type of learning behavior indicator is shown in Table 1.

What needs to be explained and distinguished is that the operation mentioned by CZ and CZZL, as shown in Table 1, refers to the command buttons on the right and low sides of geekcad, as well as the operation of mouse and keyboard light. MLZL and MLCZ refer to the 3D interactive operations on the top of geekcad, such as graphic selection, stretching, rotation, scaling.

Table 1. 3D design learning behavior indicators

Indicator	Explanation
Number of operations (CZ)	Accumulated operation times of mouse, keyboard, platform auxiliary function, etc
Number of operation types (CZZL)	Accumulated operation types of mouse, keyboard, platform auxiliary function, etc
Login duration (DLSC)	The time interval from the first operation to the last operation
Number of commands types (MLZL)	Cumulative number of types of 3D design commands such as rotate, stretch, align and crop
Number of command operations (MLCS)	Cumulative usage of 3D design commands such as rotate, stretch, align and crop

Interaction duration (JHSC)	The number of minutes in which the number of operations of 3D design command is greater than zero
Maximum operations per minute (ZDZ)	Maximum operations per minute
Maximum time interval (ZDT)	Maximum duration of zero operations per minute. Long time interval is considered as non learning state.
Efficient interaction time (YXZ)	The cumulative time that the number of clicks per minute exceeds the average number of clicks in the class

4. Result

4.1. Difference analysis of 3D design operation behavior in different situations

4.1.1. The influence of 3D model on students' 3D design

The full score of students' mental rotation ability test is 24. The pretest scores of the experimental group ranged from 2 to 18 points (M = 8.33, SD = 3.14, median = 8), while the pretest scores of the control group ranged from 0 to 18 points (M = 8.54, SD = 3.89, median = 9). By independent sample t-test, there was no difference between the experimental group and the control group (p = .085 > .05).

The pretest and posttest of each group were analyzed. As shown in Table 2, there were significant differences between the pretest and posttest of each group. Whether the experimental group or the control group, the students' spatial ability has been significantly improved after using 3D design.

Table 2. Pre- and post- test analysis of experimental group and control group

Group	Pre-test		Post-test		*t*-test		
	M	*SD*	*M*	*SD*	*MD*	*t*	*p*
Control group	8.54	3.89	12.74	5.42	-4.20	-7.55	.000**
Experimental group	8.33	3.14	13.92	4.85	-5.40	-7.38	.000**

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

After analyzing the posttest and promotion value of the two groups, it is found that, as shown in Table 3, there is no difference in the posttest value between the experimental group and the control group, that is, there is still no significant difference in the spatial ability of the two groups after the experiment. However, the improvement of spatial ability in the experimental group was significantly higher than that in the control group (p = .026 < .05).

Table 3. Comparative analysis of experimental group and control group by pre-test and post-test

Indicator	Control group (N = 46)		Experimental group (N = 51)		*t*-test		
	M	*SD*	*M*	*SD*	*MD*	*t*	*p*
Pre-test	8.54	3.89	8.33	3.14	0.21	0.291	.085
Post-test	12.74	5.42	13.92	4.85	-1.182	-1.13	.309
Improvement	4.20	3.77	5.59	5.40	-1.393	-1.46	.026*

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

Table 4. Comparative analysis of experimental group and control group by learning behavior indicators

Indicator	Control group (N = 46)		Experimental group (N = 51)		*t*-test		
	M	*SD*	*M*	*SD*	*MD*	*t*	*p*
CZ	294.09	83.905	301.08	86.286	-6.991	-0.404	0.687
CZZL	8.13	2.613	10.43	2.532	-2.301	-4.402	0.000**
DLSC	30.13	14.896	37.27	16.096	-7.144	-2.261	0.026*
MLZL	7.89	2.601	8.84	2.453	-0.952	-1.855	0.067
MLCS	68.11	42.536	81.24	35.334	-13.127	-1.659	0.100
JHSC	16.78	7.357	20.82	7.326	-4.041	-2.707	0.008**
ZDZ	81.91	37.457	63.67	36.773	18.246	2.419	0.017*
ZDT	8.67	9.825	10.47	10.542	-1.797	-0.866	0.389
YXZ	5.28	2.605	5.84	2.292	-0.561	-1.127	0.262

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

Self-directed learning (Chou, 2013), interest (Scarborough & Dobrich, 1994) and other characteristics will affect the quality and quantity of students' participation in space tasks, and then affect the growth of space ability. It is a feasible way to explore the difference of ability promotion from the perspective of behavior. It can be seen from table 4 that CZZL, DLSC and JHSC of the experimental group are significantly higher than those of the control group. However, the ZDZ of the control group was significantly higher than that of the experimental group.

4.1.2. Performance differences of students with different initial spatial abilities when using 3D models for 3D design

Similar to the first round experiment, students' spatial ability has been significantly improved after using 3D design. We will focus on the differences of learning performance among different types of students. There are many ways to distinguish high and low ability students. The median score (Boucheix & Schneider, 2009), the average score (Hu et al., 2017), and 50% of the total score (Hegarty & Steinhoff, 1997) are the three common dividing points. In this study, we choose the median of students' mental rotation test pretest score as the cut-off point. Students whose pre-test score is less than or equal to 8 will be marked as low spatial ability students, and students whose pre-test score is higher than 8 will be marked as high spatial ability students.

Through the homogeneity test of variance, we found that the variance between the high spatial ability group and the low spatial ability group was equal. Furthermore, independent sample *t*-test was performed for pretest, posttest and promotion values. As shown in Table 5, we find that there is a significant difference between the pre-test and post-test data of students' spatial ability. That is to say, the score of mental rotation test of high spatial ability students is significantly higher than that of low spatial ability students before and after training. But there is no difference between the two groups, that is, high spatial ability students and low spatial ability students' spatial ability has made the same improvement.

Table 5. Spatial ability differences between the high-and low- spatial ability students

Indicator	High spatial ability ($N = 49$)		Low spatial ability ($N = 39$)		*t*-test		
	M	*SD*	*M*	*SD*	*MD*	*t*	*p*
Pre-test	5.82	2.01	11.03	2.35	5.21	11.03	$.00^{**}$
Post-test	11.47	4.85	16.95	3.44	5.48	6.19	$.00^{**}$
Improvement	5.65	4.61	5.92	3.30	0.27	0.32	.76

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

A total of 32,657 pieces of data were collected in this experiment. The data generated by each student ranged from 199 pieces to 857 pieces ($M = 371$, $SD = 104.71$, median = 359). As shown in Table 6, the operation types, command types, command usage times, interaction time, maximum time interval and maximum operation times of high spatial ability students are slightly higher than those of low spatial ability students. But there is no significant difference between the two kinds of students. However, it is worth noting that the effective interaction time of low spatial ability students is slightly higher than that of high spatial ability students.

Table 6. Behavioral differences between the high-and low- spatial ability students

Indicator	Low spatial ability ($N = 39$)		High spatial ability ($N = 49$)		*t*-test		
	M	*SD*	*M*	*SD*	*MD*	*t*	*p*
CZ	291.43	74.850	301.59	94.903	10.161	.562	.576
CZZL	9.41	2.879	9.49	2.846	.079	.129	.898
DLSC	30.80	14.947	37.15	16.496	6.358	1.893	.062
MLZL	8.08	2.465	8.85	2.729	.765	1.378	.172
MLCS	72.47	37.546	78.56	41.025	6.095	.726	.470
JHSC	18.31	6.941	20.18	8.571	1.873	1.133	.260
ZDZ	68.67	38.107	73.21	37.321	4.532	.559	.577
ZDT	7.78	9.601	10.41	9.563	2.635	1.281	.204
YXZ	5.63	1.997	5.51	2.910	-.120	-.229	.820

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

4.2. Behavior clustering and ability improvement

4.2.1. Cluster analysis of 3D design behavior

Because the student behavior data includes time, times and other data of different dimensions, and there are extreme values between the student behavior data. We first standardize the data, and then use k-means algorithm to cluster the behavior data of students, and get five types of students, as shown in Figure 3.

Cluster5: excellent students. A total of 3 students, accounting for 3.41% of the total number of students. In addition to the maximum number of clicks and the maximum time interval, all the data such as the type of operation, the number of operations and the length of login are the highest among the students, and this type of students are the best students to participate in learning.

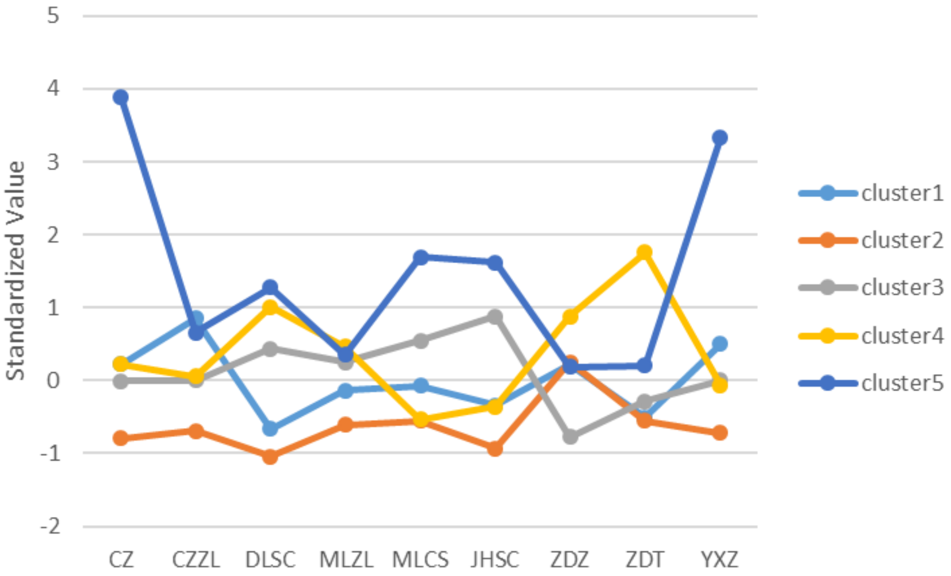
Cluster3: ordinary students. A total of 31 people, accounting for 35.23% of the total number. All the data are in the middle of all the students, and are basically above the average level.

Cluster2: risk students. A total of 23 people, accounting for 26.14% of the total number. In addition to the maximum number of clicks, all the data such as the type of operation, the number of operations and the length of login are the worst among the students, and there is a huge gap with the average level. This type of students is the worst in learning participation.

Cluster1: tasters. A total of 15 people, accounting for 17.05% of the total number. The operation types and effective interaction time are above the average level, but the operation times and command types are below the average level, and the login time is relatively short. This type of students spend more time using some platform operations and 3D design operations in less login time.

Cluster4: task quitting. A total of 16 people, accounting for 18.20% of the total number. Although online for a long time, the maximum time interval is particularly large, that is, the middle of a particularly long time did not participate in 3D design. It is considered that the students once gave up their study in the middle of the course.

Figure 3. Comparison of five types of students' operation behavior



4.2.2. Differences of spatial ability among different types of students

As shown in Table 7, there are no significant difference in the pre-test (p = .145 > .05) and post test (p = .285 > .05) of the five types of students. After analyzing the five types of students, we can find that there is no difference between the five types of students in the post test. But in the pretest, there is significant difference between cluster1 and cluster5. On the whole, there is no significant difference in improvement of the five types of students (p = .053 > .05), but there was significant difference between cluster 2 and cluster 1, cluster 3, Cluster 4.

Table 7. Differences of spatial ability among different types of students

Indicator	Types	*M*	*SD*	*F*	*Sig.*
Pre-test	Cluster 1	6.73	2.915	1.757	0.145
	Cluster 2	8.39	4.076		
	Cluster 3	7.90	3.070		
	Cluster 4	8.81	2.857		
	Cluster 5	11.67	3.512		
Post-test	Cluster 1	14.33	4.865	1.279	0.285
	Cluster 2	12.17	5.622		
	Cluster 3	13.97	4.902		
	Cluster 4	15.25	4.754		
	Cluster 5	17.00	2.646		
Improvement	Cluster 1	7.60	4.067	2.444	0.053
	Cluster 2	3.78	3.357		
	Cluster 3	6.06	4.676		
	Cluster 4	6.44	3.140		
	Cluster 5	5.33	1.155		

5. Discussion

5.1. The influence of 3D model on students' spatial ability

Through paired sample *t*-test of students' mental rotation ability before and after the test, it is found that no matter what kind of teaching intervention materials are provided, 3D design can significantly improve students' spatial ability. That is, with the help of 3D design technology, students' spatial ability has been significantly improved, which is consistent with the research of Uttal et al., 2013. The improvement value of spatial ability of the experimental group was significantly higher than that of the control group, which seems to indicate that compared with paper materials, providing 3D model in the process of 3D design can improve students' spatial ability more effectively. However, it is worth noting that there is no difference between the pre-test and posttest of spatial ability between the experimental group and the control group. From the final results, we cannot conclude that 3D model intervention can improve students' spatial ability more effectively. 3D design operation data can let us interpret this phenomenon more accurately.

DLSC, CZZL and JHSC of the experimental group were significantly higher than those of the control group, while ZDZ of the control group was significantly higher than that of the experimental group. 3D model teaching intervention can make students spend more time on the platform and use more commands unrelated to 3D design for 3D design tasks, so that the 3D model education intervention training achieved significantly higher than the paper material education intervention value. However, although the 3D interaction time is long, there is no difference in the types of 3D design operations between the experimental group and the control group, so there is no difference in the post test of spatial ability between the experimental group and the control group. Similar to Scarborough and Dobrich (1994), 3D model intervention may enhance students' interest in learning, make students spend more time, carry out more operations, and achieve greater value of spatial ability improvement. However, because the novelty of 3D model will also enhance students' interest in learning, in the long-term experimental environment, students' novelty or interest in learning may change. We also need to verify students' differences in different educational intervention situations according to different task difficulty.

5.2. The differences of spatial ability training among students with different initial levels of spatial ability

Similarly, through the second round of experiments, we also found that using 3D models to carry out 3D design can significantly improve students' spatial ability. Through paid matched samples *t*-test on students' mental rotation before and after test scores, we found that 3D design can significantly improve students' spatial ability. Furthermore, we find that after 3D design, high spatial ability students and low spatial ability students have the same improvement. This finding is different from the ability-as-enhancer hypothesis and ability-as-compensator hypothesis. It is also possible that 3D design enables students to operate 3D objects intuitively, which reduces students' cognitive load and makes no difference between students.

However, after 3D design training, the mental rotation performance of high spatial ability students is still better than that of low spatial ability students. However, the experiment we set up is relatively simple, and ignores the individual differences of motivation and interest besides the differences of learners' abilities, and does not use a variety of educational interventions to ensure the diversity of learning environment design (Höffler, 2010). This conclusion needs more extensive verification. In the 3D design training, teachers should get rid of the stereotype in spatial ability and carry out teaching activities equally.

Therefore, our research no longer focuses on the performance of students with different initial spatial abilities in different teaching situations, but explores the performance differences of students with different initial spatial abilities from the perspective of learning analysis. Due to the short research cycle and low difficulty of the task, the research results need to be further verified in order to promote. In addition, this study uses 3D design to train students' spatial ability, but still uses traditional MRT to measure students' spatial ability. Although most of the existing studies also use this method (Koesa & Karakus, 2018), the effectiveness of the measurement results needs to be further verified.

5.3. Behavior differences of students with different spatial levels

At present, there are many studies on building block activities, computer-aided design, sketch, 3D modeling and other activities on the cultivation of spatial ability, but there is no research on the effect of spatial ability cultivation based on students' behavior data. In the field of online learning, many scholars regard interaction as the most important part of all learning environments (Woo & Reeves, 2007). Many scholars extract students' behavior variables from the system log data, and explore the behavior variables to predict students' performance (Macfadyen & Dawson, 2010; Morris et al., 2005). Similar to this study, Sherman and Martin proposed a method to capture student app inventor project snapshot and explore student development behavior (Sherman & Martin, 2015). Filvà collects the data generated in students' scratch interaction, detects students' behavior patterns, and supports teachers to provide implementation quality feedback (Filvà et al., 2019). We collected the data of students' operation times, login time, operation time interval in the process of 3D design, and analyzed the behavior data of high spatial ability and low spatial ability. We observed that high spatial ability students used more command types and operated more times in a longer time than low spatial ability students. Although most of the behavior data of high spatial ability is higher than that of low spatial ability, there is no significant difference between the two types of students.

The results of this study are not consistent with either the ability-as-enhancer hypothesis or the ability-as-compensator hypothesis. For one thing, ability-as-enhancer hypothesis and the ability-as-compensator hypothesis are mainly based on the cognitive load theory. And 3D modeling provides students with intuitive 3D space operation experience, which does not require students to convert abstract two-dimensional or text data into 3D objects for further operation, and will not cause cognitive overload of students with low spatial ability. For another, it is also possible that the task of this study is relatively simple, the difficulty of the task and the sense of achievement of the task can not meet the learning desire of the high spatial learners, which makes the 3D object operation become a low desire learning activity, and produce similar learning results and behaviors with the low spatial learners. The results also need to carry out a longer period of research in the complex teaching situation to verify.

5.4. The improvement of spatial ability of different types of students

Referring to the interaction behavior in the online learning environment, students' 3D modeling behavior is also related to students' ability to use specific learning tools and find the right information. These abilities will also affect students' 3D design operation, and further affect the cultivation of spatial ability (Hillman et al., 1994; Lust et al., 2012). In addition, learning situation, external motivation, instructional design and task setting also affect students' 3D modeling enthusiasm and their 3D modeling behavior. Regardless of students' internal and external motivation, it will be reflected in the 3D modeling behavior and affect the improvement of spatial ability. Therefore, only using the initial spatial ability to analyze students' 3D modeling behavior, there are still many uncertainties. In order to further explore the improvement of different students' spatial ability, we use k-means algorithm to cluster students' 3D modeling behavior, and get five types of students.

Novice designers often test their designs through trial and error, and experienced designers use different testing strategies according to their experience (Ahmed et al., 2003). Although we cannot fully refer to students' 3D modeling behavior data to evaluate students' spatial ability, behavioral data can provide a reference for the cultivation and improvement of students' spatial ability. Although there are only three students in cluster5, it is the most ideal type of learning. Cluster4 is in the state of not learning for a long time. It is possible that the students have completed the course task in a short time and have a long idle period. It is also possible that the students have not seriously participated in the course activities. Both cluster1 and cluster2 participate in 3D modeling, but both cluster1 and cluster2 are of poor participation type. Compared with cluster2, cluster1 tries more 3D modeling operations and has a longer interaction time. Cluster1 and cluster2 may be that the students have completed the course task in a short time, and then they are not interested in participating in the course. Cluster3 belongs to ordinary students, all behaviors are above the average, and the number of students is the largest.

From the perspective of pretest, there are some differences in 3D modeling behavior between students with low spatial ability and students with high spatial ability. Students with particularly low spatial ability may encounter great difficulties in the process of 3D modeling. Therefore, although they are more interested in 3D design, have tried the platform functions, and have relatively long interaction time, they do not continuously participate in the course activities and become tasters. Students with high spatial ability are more interested in 3D design and continue to participate in the course tasks. Only the students with very low spatial ability and very high spatial ability showed differences in 3D modeling behavior. There is no difference in the post test and spatial ability improvement between the two kinds of students. As long as students participate in 3D modeling, their spatial ability has been trained and improved. Although students' 3D modeling behavior is different due to internal and external motivation, it can be considered that students have similar spatial ability after the course. From the perspective of promotion, the promotion value of spatial ability of cluster2 is significantly less than that of cluster1, cluster3 and cluster4, but there is no difference with that of cluster5. Therefore, teachers should encourage and guide students to actively participate in 3D design, give more guidance to students with low spatial ability, and pay attention to personalized feedback of students with high spatial ability. Although we often choose a few students as representatives to praise in teaching practice, the relevant conclusion can only be a hypothesis, because the number of cluster5 is too small. More extensive research is needed to promote the experimental results.

6. Conclusion

The results of this exploratory study verify the following points: (1) On the whole, 3D design can improve students' spatial ability, regardless of their initial spatial ability. (2) 3D model can enhance students' interest in learning and encourage them to spend more time on 3D design. (3) Students with different spatial ability levels still have significant differences in their spatial ability after using 3D design, but there is no significant difference in the improvement value of spatial ability among students with different spatial ability levels.

We also found that: (1) Spending more time and exploring more 3D design functions can improve students' space ability more effectively. Teachers should take the initiative to set up educational intervention to enhance students' learning interest and motivation, so as to improve students' learning effect. However, in the long-term process of education, the impact of educational intervention on students' interest and the persistence of students' spatial ability need to be further explored. (2) The 3D modeling behavior data of high spatial ability students is better than that of slightly low spatial ability students. High spatial ability students spend more time using more command types and doing more operations, but low spatial ability students have longer effective interaction time. However, it should be emphasized that there is no significant difference in 3D modeling behavior data among students with different spatial ability levels. (3) There are many types of students' behavior data in 3D design, including excellent students, ordinary students, risk students, tasters and task quitting. The students with great differences in initial spatial ability also have differences in their operation behavior, but most of them have no differences in their operation behavior. There was no significant difference in the post test value of spatial ability among different types of students, but the students with the worst participation achieved the least improvement in spatial ability.

In order to explore the differences of students' 3D modeling behavior, we need to extract computable key indicators, and provide personalized guidance and timely feedback to students. Although the behavior data of 3D design in this study cannot predict the improvement of students' spatial ability. There are also other ways to predict the effect or promotion of students' spatial ability according to the initial level of spatial ability. For example, Xiao and Zhang (2021) through two years of continuous research, found that interest in space activities can significantly predict the development of spatial ability, but has nothing to do with the initial spatial ability. Turgut (2015) developed the Spatial Ability Self-Report Scale (SASRS) to evaluate the spatial ability of college students. Despite, the SASRS is not suitable for k1k8 students. However, it can be used for high school students. The automatic tracking, collection and storage of 3D design behavior data can avoid the Pygmalion effect and Hawthorne effect, which are often encountered in empirical research, and provide a path for the study of spatial ability improvement differences.

In short, we verify the spatial ability growth of students with different initial spatial abilities in 3D design, and explore the spatial ability growth of students with different initial spatial abilities and different types of students through behavioral data. It is a limitation of this study to choose mental rotation measurement as the basis of students' spatial ability. At present, there is no unified definition of spatial ability, and there is no perfect measurement scale of each part and comprehensive use experience. Mental rotation and other spatial ability scale can only measure part of the content of spatial ability. Moreover, spatial ability is trained by 3D design, but the typical mental rotation scale is only a simple graphic test, and the difference between measurement methods and training methods will also affect the credibility of the measurement results. Moreover, the results of mental rotation measurement are also affected by students' speech ability and logical reasoning ability. The choice of research objects is another limitation. Although the purpose of this study is to explore the growth of spatial ability of students with different initial spatial ability and different learning types, we can not extend the results to all middle school students because we do not consider the prior knowledge of students and the difficulty of this 3D design task to be low. In addition, although learning behavior data is the result of the comprehensive influence of students' internal and external motivation, different teaching strategies, different task scenarios and different operation tools may also have different effects on students' behavior data, and students' external motivation is still an important factor affecting the research results.

Acknowledgement

This study is supported by the Humanity and Social Science foundation from the Ministry of Education of China (grant number 20YJA880057) and Excellent Doctoral Students' academic innovation ability promotion project at East China Normal University (grant number YBNLTS2020-004).

Reference

Ahmed, S., Wallace, K. M., & Blessing, L. T. (2003). Understanding the differences between how novice and experienced designers approach design tasks. *Research in Engineering Design, 14*(1), 1-11.

Al-Ahmari, A., Ameen, W., Abidi, M. H., & Mian, S. H. (2018). Evaluation of 3D printing approach for manual assembly training. *International Journal of Industrial Ergonomics, 66*, 57-62.

Baenninger, M., & Newcombe, N. (1989). The Role of experience in spatial test performance: A Meta-analysis. *Sex Roles, 20*(5), 327-344.

Barber, S. R., Kozin, E. D., Dedmon, M., Lin, B. M., & Lee, D. J. (2016). 3D-printed pediatric endoscopic ear surgery simulator for surgical training. *International Journal of Pediatric Otorhinolaryngology, 90*, 113-118.

Blikstein, P., Kabayadondo, Z., Martin, A., & Fields, D. (2017). An Assessment instrument of technological literacies in makerspaces and FabLabs. *Journal of Engineering Education, 106*(1), 149-175.

Boucheix, J. M., & Schneider, E. (2009). Static and animated presentations in learning dynamic mechanical systems. *Learning & Instruction, 19*(2), 112-127.

Carroll, J. B. (1993). *Human cognitive abilities: A Survey of factor-analytic studies*. Cambridge University Press.

Chien, Y.-H., & Chu, P.-Y. (2018). The Different learning outcomes of high school and college students on a 3D-printing STEAM engineering design curriculum. *International Journal of Science and Mathematics Education, 16*(6), 1047-1064.

Chou, P.-N. (2013). Students' perceptions of success in the online graduate-level classes: A Self-directed learning perspective. *Contemporary Issues in Education Research, 6*(1), 115-122.

Contreras, M. J., Rubio, V. J., Peña, D., Colom, R., & Santacreu, J. (2007). Sex differences in dynamic spatial ability: The Unsolved question of performance factors. *Memory & Cognition, 35*(2), 297-303.

Deno, J. A. (1995). The Relationship of previous experiences to spatial visualization ability. *Engineering Design Graphics Journal, 59*(3), 5-17.

Eisenberg, M. (2013). 3D printing for children: What to build next? *International Journal of Child-Computer Interaction, 1*(1), 7–13. https://doi.org/10.1016/j.ijcci.2012.08.004

Filvà, D. A., Forment, M. A., García-Peñalvo, F. J., Escudero, D. F., & Casañ, M. J. (2019). Clickstream for learning analytics to assess students' behavior with Scratch. *Future Generation Computer Systems, 93*, 673-686.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research*, 74(1), 59-109. https://doi.org/10.3102/00346543074001059

Gerson, H. B., Sorby, S. A., Wysocki, A., & Baartmans, B. J. (2001). The Development and assessment of multimedia software for improving 3-D spatial visualization skills. *Computer Applications in Engineering Education, 9*(2), 105-113.

Hegarty, M. , & Sims, V. K. (1994). Individual differences in mental animation during mechanical reasoning. *Memory & Cognition*, 22(4), 411-430.

Höffler, T. N. (2010). Spatial ability: Its influence on learning with visualizations—A Meta-analytic review. *Educational Psychology Review, 22*(3), 245-269.

Höffler, T. N., & Leutner, D. (2011). The Role of spatial ability in learning from instructional animations–Evidence for an ability-as-compensator hypothesis. *Computers in Human Behavior, 27*(1), 209-216.

Hays, T. A. (1996). Spatial abilities and the effects of computer animation on short-term and long-term comprehension. *Journal of Educational Computing Research, 14*(2), 139-155. https://doi.org/10.2190/60Y9-BQG9-80HX-UEML

Heckman, J. J., & Masterov, D. V. (2007). The Productivity argument for investing in young children. *Applied Economic Perspectives and Policy, 29*(3), 446-493.

Hillman, D. C., Willis, D. J., & Gunawardena, C. N. (1994). Learner-interface interaction in distance education: An Extension of contemporary models and strategies for practitioners. *American Journal of Distance Education, 8*(2), 30-42.

Hu, Y., Wu, B., & Gu, X. (2017). An Eye tracking study of high-and low-performing students in solving interactive and analytical problems. *Educational Technology & Society, 20*(4), 300-311.

Huang, T. C., & Lin, C. Y. (2017). From 3D modeling to 3D printing: Development of a differentiated spatial ability teaching model. *Telematics & Informatics, 34*(2), 604-613.

Huk, T. (2010). Who benefits from learning with 3D models? the case of spatial ability. *Journal of Computer Assisted Learning, 22*(6), 392-404.

Kim, D., Park, Y., Yoon, M., & Jo, I. H. (2016). Toward evidence-based learning analytics: Using proxy variables to improve asynchronous online discussion environments. *The Internet and Higher Education*, 30, 30-43. https://doi.org/10.1016/j.iheduc.2016.03.002.

Kleij, F. M. V. D., Eggen, T. J. H. M., Timmers, C. F., & Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education, 58*(1), 263-272.

Koesa, T., & Karakus, F. (2018). The Effects of computer-aided design software on engineering students' spatial visualisation skills. *European Journal of Engineering Education, 43*(2), 296-308.

Larson, G. E. (1996). Mental rotation of static and dynamic figures. *Perception & Psychophysics, 58*(1), 153-159.

Lee, A. L., & Wong, K. W. (2014). Learning with desktop virtual reality: Low spatial ability learners are more positively affected. *Computers & Education, 79*, 49-58. https://doi.org/10.1016/j.compedu.2014.07.010

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A Meta-analysis. *Child Development, 56*(6), 1479-1498.

Lubinski, D. (2010). Spatial ability and STEM: A Sleeping giant for talent identification and development. *Personality & Individual Differences, 49*(4), 344-351.

Lust, G., Collazo, N. A. J., Elen, J., & Clarebout, G. (2012). Content management systems: Enriched learning opportunities for all? *Computers in Human Behavior, 28*(3), 795-808.

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A Proof of concept. *Computers & Education, 54*(2), 588-599.

Hegarty, M., & Steinhoff, K. (1997). Individual differences in use of diagrams as external memory in mechanical reasoning. *Learning & Individual Differences, 9*(1), 19-42.

Masters, M. S. (1998). The Gender difference on the Mental Rotations test is not due to performance factors. *Memory & Cognition, 26*(3), 444-448.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of educational psychology, 86*(3), 389–401. https://doi.org/10.1037/0022-0663.86.3.389

Mcleod, D. B., Cronbach, L. J., & Snow, R. E. (1977). Aptitudes and instructional methods: A Handbook for research on interactions. *Journal for Research in Mathematics Education, 9*(5), 390-395.

Meijer, F., & Broek, E. L. V. D. (2010). Representing 3D virtual objects: Interaction between visuo-spatial ability and type of exploration. *Vision Research, 50*(6), 630-635.

Monahan, J. S., Harke, M. A., & Shelley, J. R. (2008). Computerizing the mental rotations test: Are gender differences maintained? *Behavior Research Methods, 40*(2), 422-427.

Morris, L. V., Finnegan, C., & Wu, S. S. (2005). Tracking student behavior, persistence, and achievement in online courses. *The Internet and Higher Education, 8*(3), 221-231.

Ortner, T. M., & Sieverding, M. (2008). Where are the gender differences? Male priming boosts spatial skills in women. *Sex Roles, 59*(3-4), 274-281.

Šafhalter, A., Glodež, S., Šorgo, A., & Ploj Virtič, M. (2020). Development of spatial thinking abilities in engineering 3D modeling course aimed at lower secondary students. *International Journal of Technology and Design Education*, 1-18. https://doi.org/10.1007/s10798-020-09597-8

Scarborough, H. S., & Dobrich, W. (1994). On the efficacy of reading to preschoolers. *Developmental Review, 14*(3), 245-302.

Sharps, M. J., Price, J. L., & Williams, J. K. (2010). Spatial cognition and gender instructional and stimulus influences on mental image rotation performance. *Psychology of Women Quarterly, 18*(3), 413-425.

Sherman, M., & Martin, F. (2015). Learning analytics for the assessment of interaction with App Inventor. *The 2015 IEEE Blocks and Beyond Workshop (Blocks and Beyond)* (pp. 13-14). IEEE. https://doi.org/10.1109/BLOCKS.2015.7368992

Sims, V. K., & Mayer, R. E. (2002). Domain specificity of spatial expertise: The Case of video game players. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 16*(1), 97-115.

Snyder, T. J., Weislogel, M., Moeck, P., Stone-Sundberg, J., & Graft, J. (2014). 3D systems technology overview and new applications in manufacturing, engineering, science, and education. *3D Printing and Additive Manufacturing, 1*(3), 169-176.

Sorby, S., Casey, B., Veurink, N., & Dulaney, A. (2013). The Role of spatial training in improving spatial and calculus performance in engineering students. *Learning & Individual Differences, 26*, 20-29.

Turgut, M. (2015). Development of the spatial ability self-report scale (SASRS): Reliability and validity studies. *Quality & Quantity, 49*(5), 1997-2014.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The Malleability of spatial skills: a meta-analysis of training studies. *Psychological bulletin*, 139(2), 352. https://doi.org/10.1037/a0028446

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*(2), 599-604.

Voyer, D., Rodgers, M. A., & Mccormick, P. A. (2004). Timing conditions and the magnitude of gender differences on the Mental Rotations Test. *Memory & cognition, 32*(1), 72-82.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A Meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*(2), 250-270.

Whitelock, D. (2009). e-Assessment: Developing new dialogues for the digital age. *British Journal of Educational Technology, 40*(2), 199-202.

Woo, Y., & Reeves, T. C. (2007). Meaningful interaction in web-based learning: A Social constructivist interpretation. *The Internet and Higher Education, 10*(1), 15-25.

Wu, Y.C., Liao, W.H., Liu, C.Y., Li, T.Y., & Chi, M.T. (2018). Evaluation of student's 3D modeling capability based on model completeness and usage pattern in K-12 classrooms. *The 2018 IEEE 18th international conference on advanced learning technologies (ICALT)* (pp. 244-248). IEEE. https://doi.org/10.1109/ICALT.2018.00063

Wu, Y., Guo, S., & Zhu, L. (2020). Design and implementation of data collection mechanism for 3D design course based on xAPI standard. *Interactive Learning Environments, 28*(5), 602-619.

Xiao, N., & Zhang, X. (2021). Interest in spatial activities predicts young children's spatial ability development: A Two-year longitudinal study. *Contemporary Educational Psychology*, 101943. https://doi.org/10.1016/j.cedpsych.2021.101943

Yong, B., Jiang, X., Lin, J., Sun, G., & Zhou, Q. (2022). Online Practical Deep Learning Education: Using Collective Intelligence from a Resource Sharing Perspective. *Educational Technology & Society*, *25* (1), 193-204.

Online Practical Deep Learning Education: Using Collective Intelligence from a Resource Sharing Perspective

Binbin Yong[1], Xuetao Jiang[1], Jiayin Lin[2], Geng Sun[3] and Qingguo Zhou[1*]

[1]School of Information Science and Engineering, Lanzhou University, Lanzhou, China // [2]School of Computing and Information Technology, University of Wollongong, Wollongong, NSW, Australia // [3]Vermilion Cloud, Surrey Hills, NSW, Australia // yongbb@lzu.edu.cn // jiangxt18@lzu.edu.cn // jl461@uowmail.edu.au // gs147@uowmail.edu.au // zhouqg@lzu.edu.cn

[*]Corresponding author

ABSTRACT: Deep learning (DL), as the core technology of artificial intelligence (AI), has been extensively researched in the past decades. However, practical DL education needs large marked datasets and computing resources, which is generally not easy for students at school. Therefore, due to training datasets and computing resources restrictions, it is still challenging to popularize DL education in colleges and universities. This paper considers solving this problem by collective intelligence from a resource sharing perspective. In DL, dataset marking and model training both require high workforce and computing power, which may implement through a resource sharing mechanism using collective intelligence. As a test, we have designed a DL education scheme based on collective intelligence under the background of artistic creation to collect teaching materials for DL education. Also, we elaborate on the detailed methods of sharing mechanisms in this article and discuss some related problems to verify this shared learning mechanism.

Keywords: Deep learning education, Datasets and computing resources, Collective intelligence, Resource sharing perspective

1. Introduction

Artificial Intelligence (AI) has attracted the attention of many researchers since the invention of computers. A lot of work has been done to endow machine intelligence. However, only in recent years, with the development of deep learning (DL) technology (Lecun et al., 2015), AI has made significant breakthroughs in theories and applications. It was proved that a multi-layer perceptron (MLP) could simulate any function with a lot of computation for training the MLP model (Pinkus, 1999). The backpropagation (BP) algorithm (Hameed et al., 2016) was then applied to train a shallow neural network model (equal to MLP) in the 1980s when the personal computer (PC) was invented, and the computing power was greatly improved. Then, neural network technologies had remained stagnant until the mid-2000s because of the small samples and low computing power. In fact, the support vector machine (SVM) (Zhi et al., 2018) model was widely studied and used in this period due to its high accuracy for small samples.

With the development of mobile Internet technology, data has shown a trend of explosive growth since the beginning of this century, which has brought the era of big data (Daniel, 2015). One point of big data is that knowledge is no longer a concise statement or a formula. It is indeed stored in massive data. Therefore, a type of model that can learn the internal statistical characteristics of big data is needed.

Meanwhile, the computing performance of computers increases exponentially. Both big data and high computing performance two factors triggered the revolution of DL technologies. According to the law of large numbers, the empirical risk of a forecasting model will tend to be expected risk when the number of training samples tends to be infinite. Therefore, with the advent of big data, more data brings more accurate predictions. On the other hand, according to the basic theory of neural networks, three layers neural network can simulate any functions with enough hidden nodes. In 2012, Krizhevsky et al. (2012) proved deep neural networks' strong patterns recognition performance combined with big data. Later research indicated that a deeper neural network could be fully utilized to mine the data rules and make predictions. A Deeper neural network can bring more substantial capabilities, feature extraction, and feature learning capabilities (Ayinde et al., 2019). Meanwhile, a deeper neural network also needs more computing resources to train, which is still difficult for ordinary college students. Overall, the practical learning of DL is mainly based on two essential conditions: large amounts of training datasets and adequate computing resources.

Nowadays, typical DL applications, such as image recognition (He et al., 2015), speech recognition (Hinton et al., 2012), natural language processing (Tingting & Mengyu, 2019), autonomous vehicle (Ye et al., 2018), and robot (Chao et al., 2019) are widely researched and developed (Lo & Shu, 2005). A decade ago, Welham (2008) discussed the difficulties of using AI in education. The paper showed several issues that slowed down the pace of AI entering the field of education. However, due to the rapid development of DL technologies, the environment has changed a lot. DL technologies have revolutionized research methods in many fields, and it also changed the content of the AI-related education curriculum. In 2019, 35 universities in China were first approved to add AI specialty for undergraduates (China Daily [CD], 2019). AI used to be an elective major at the undergraduate level in China. Current AI courses and textbooks are generally too old to meet teaching needs. Therefore, new materials and procedures containing the latest DL contents need to be developed for undergraduates. As the critical issues in DL, dataset and computation problems (Krizhevsky et al., 2012) for DL practice learning need to be researched for available teaching schemes.

Aimed at sharing Chinese artwork by DL course, we chose Dunhuang data as our material. Dunhuang murals are the representatives of ancient Chinese grotto murals. Although these murals have had a history of thousands of years, they have shown different styles in painting images and color forms. They utilized a small number of colors to paint pictures, which did not make people feel monotonous. In contrast, they show a unique gorgeous. Dunhuang frescoes reflect ancient China's culture and art style, which can be well used for reference by today's art design. Dunhuang's mural images show different characteristics in the various dynasties. They can also be divided into different categories according to their contents, such as flying apsaras, landscape painting, Buddhistic stories, etc.

In this paper, we propose schemes to address these DL education problems by using collective intelligence from a resource sharing perspective. Based on the background of artistic creation, we design a scheme to collect teaching materials for the DL teaching test. Then, we found that the shared approaches taken in the design of the DL course can help address the datasets and computing challenges currently present in DL education. We also did a short DL course practice to test the learning effect. Furthermore, this paper gives the artworks designed and implemented by students in class. Also, evaluation results of the DL course are given according to the questionnaire results. It should be emphasized that we are not providing perfect course schemes that can be applied in any scenario. In fact, we present some feasible approaches in practice for general DL education. Teachers can change the DL course contents based on schemes proposed in this paper and specific teaching environments.

This paper will focus on using a practice perspective to share the datasets and computing resources. We will explore the resource sharing methods, suitable and typical DL course contents, and project practice content. Compared with previous work, our research mainly has the following contributions.

- From a resource sharing perspective, a collective intelligence scheme is designed for DL course materials.
- A concrete sample for using collective intelligence for the DL course is conducted based on the background of art creation.
- Approaches to solving the dataset and computing problems are given and discussed.
- A DL course practice is conducted to verify the DL education effect. And the course practice results and evaluation are given and analyzed.

The rest of this paper is organized as follows. Section 2 analyzes the investigations of current challenges for DL education. In Section 3, we present the coping strategies for these challenges. Section 4 present the contents of the DL course. Section 5 presents the discussion and conclusion of this study.

2. Current challenges

As discussed in the introduction section, the critical challenges in practical DL education mainly focus on the course materials, especially dataset and computation problems, which are rarely discussed in the literature. And there is little research that focuses on this.

2.1. Course challenge

For DL courses, some open courses can be found online, such as MOOC (Freitas et al., 2015; Liyanagunawardena et al., 2015; Terras & Ramsay 2015), Coursera, and Udacity (Giannakos, 2013), etc. These online courses have promoted the development of DL education. Although many courses have well-designed interaction, free courses, and optional paid certification, which enable students to learn DL courses online

without difficulty, these courses offer a little place for collaboration. It means that students learn and practice DL alone instead of studying as a group, which is very important for the development of modern DL.
Course contents need to be arranged in natural environments for DL practical education according to specific conditions, such as the differences of knowledge foundation, educational level, and professional direction. Even for college students of computer science majors, the DL course is difficult for them. Therefore, DL education needs to select and organize the course contents.

2.2. Dataset challenge

For training datasets, there have been many public datasets for DL of different purposes, such as MNIST (Deng, 2012), ImageNet (Deng et al., 2009), COCO (Lin et al., 2014), etc. However, we need to collect and mark datasets for our applications in many cases. Generally, good datasets are more critical than DL models for DL applications, and dataset marking is essential but tedious. However, many online DL courses use public datasets for experiments. In this case, learners do not know how to make their datasets for their application scenarios. Datasets making needs to solve the following problems. Firstly, we must collect interesting input data (e.g., images). We can collect the input data by a hand-made approach, but it takes a lot of time. Another commonly used way is to collect these data by crawling through the Internet. And the problem is that unwanted data may be crawled due to the uncertainty of network contents. Secondly, it is a vast project to organize and mark these collected data. There have been many ready-made marking tools, such as LabelMe, LabelImg etc. However, data marking needs lots of people and time, and hidden dangers of mislabeling exist, which may cause low accuracy of the DL model. In fact, task assignment for data marking is also a project problem, and there is a lack of a specific case for the student to referentially cooperate in data marking.

2.3. Computing challenge

The development of a neural network is consistent with the development of the computer. The growth of computing power promotes the increase in the size and complexity of neural networks. According to our test, training a simple LeNet5 model (5 layers) for MNIST digital recognition (Lecun et al., 1998) will take about one hour on Intel Core i5 CPU, and it takes only 90 seconds on Nvidia GTX 1080 GPU. But parallel computing device is too expensive for most commercial cloud servers are costly for undergraduates. They do not have adequate flexibility to be assigned to students. Colab is a cloud server platform of Google designed for DL and machine learning tasks. Colab is a cloud server platform of Google designed for DL and machine learning tasks, but such platforms may not be accessible in Chine due to the national policies.

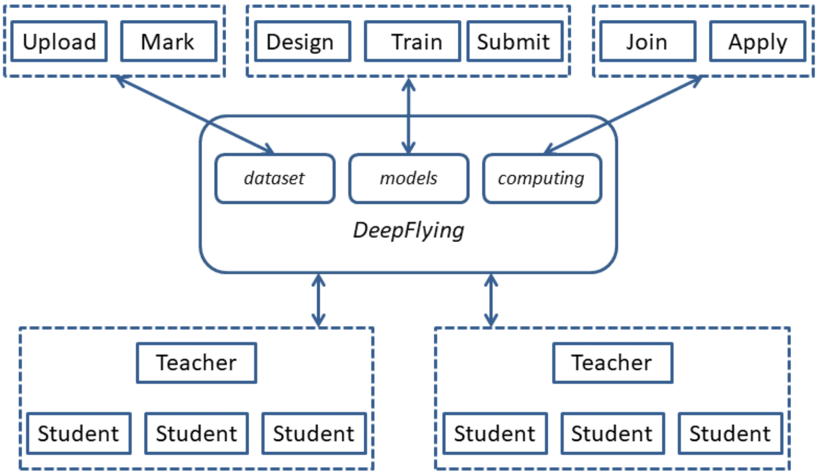
Moreover, one of the challenges is that in many universities with poor conditions (such as non-key universities), there is no condition to establish or purchase cloud service platforms for students. Generally, students have PC. And the DeepFlying (Deep learning and Flying apsaras) platform proposed in this paper has the advantage of centralizing and dispersing PC computing power to complete DL teaching practice. Therefore, the other choice is to buy components or cloud servers, which is very difficult for many students and poor universities (especially in Western China).

3. Addressing the common challenges

To address the challenges in practical DL education, we provide some solutions and test them in practice, as shown in Figure 1. In our DL course practice, we try on the Dunhuang theme. Therefore, we name our DL education platform DeepFlying, which utilizes collective intelligence to collect and share DL materials and resources. The shared resources mainly include datasets, DL models, and computation.

Teachers and students in different universities can access the DeepFlying platform to upload and mark data. Also, they can share their DL models on the DeepFlying platform and apply for computing resources for training DL models. Therefore, DL education materials, dataset making, and computing resources are essential for the DeepFlying platform.

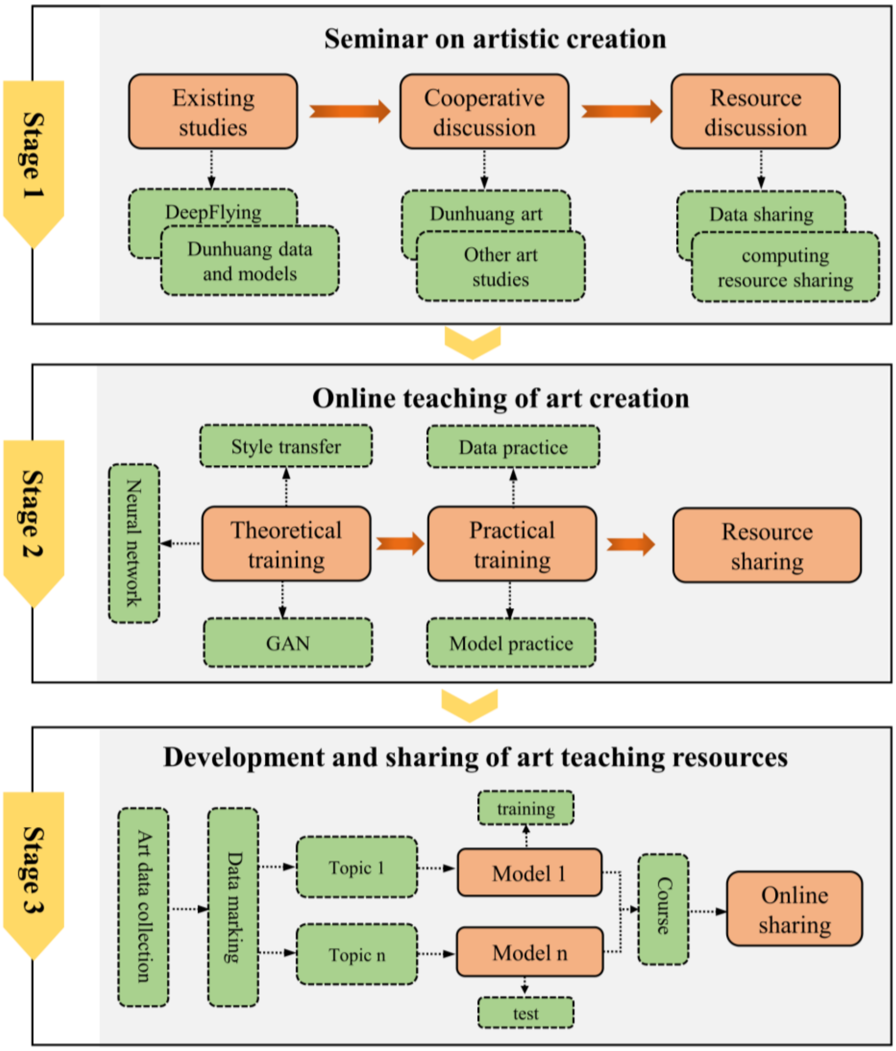
Figure 1. The schemes to address practical DL education challenges



3.1. DL education materials

To collect the ideas and teaching materials for DL education, we design a scheme based on the background of artistic creation. As shown in Figure 2, the process mainly consists of three stages. And the DL education is organized after these stages.

Figure 2. Three stages by using collective intelligence for DL education



3.1.1. Seminar on artistic creation

The first stage in the DL education is to hold a seminar on artistic creation. This seminar aims to identify topics and ideas for deep learning education among different universities. Existing studies such as the DeepFlying platform, Dunhuang dataset, and models are presented and discussed in the seminar. Next, more research ideas about this field will be put forward. These cooperative discussions may generate more teaching materials, including Dunhuang or other studies. At the end of the first phase, data and computing resource sharing methods are discussed.

3.1.2. Online teaching of art creation

The second stage in the DL education is to conduct online education on art creation. In this way, we can test the teaching materials and choose suitable teaching materials based on students' feedback. The online course includes theoretical training and practical training. In the theoretical training part, the theories of neural network (NN), style transfer neural network (STNN), and generative adversarial network (GAN) are introduced. Then, students are taught how to mark datasets, design, and train models in the practical training part. The marked dataset and trained models will be shared online.

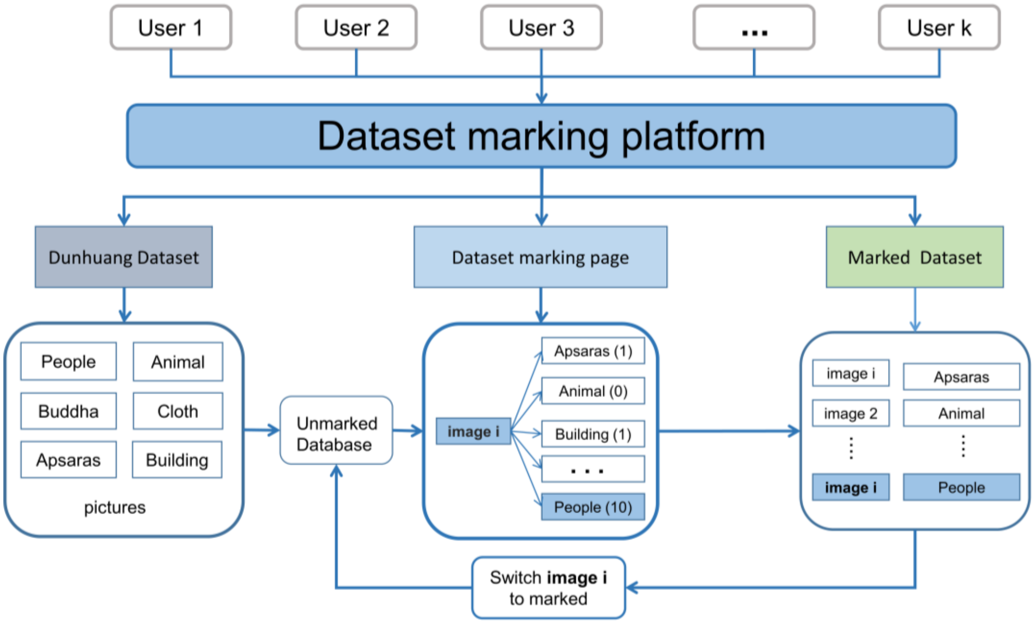
3.1.3. Development and sharing of art teaching resources

In stage 3, based on the experience of the first two stages, DL teaching on a specific artistic topic is conducted by different teachers. Students are taught to collect and share more art data in the course. Meanwhile, students finish data marking as their homework. Then, based on the open dataset, different DL topics are proposed and implemented by DL models. Students can team up to work on various DL tasks, such as each team focusing on a single classification function, and they are guided to train and test their models. At last, the materials, including data and models, are made into the open course and shared online.

3.2. Dataset making

Data collecting and marking are important for DL education. To enable students to collaborate in collecting and marking required datasets, we propose and design an online data collecting and marking module in the DeepFlying platform, denoted as dataset marking module, as shown in Figure 3.

Figure 3. The dataset marking module in DeepFlying



Each student needs to register as a user in the marking module, and students collect the dataset according to their homework. The dataset collection task is evenly assigned to students who attend the DL courses, and each student needs to collect and upload a portion of the dataset. Then, each student needs to register an account and mark parts of the dataset online, different from the data they collected. The marked data is randomly selected from the whole dataset. Each data is marked at least ten times, more trials would be applied if the most marked category has less count, and the top marked category is selected as the final category of the data. In this way, data will only be marked when students reach a consensus, preventing students from cheating when marking data or obtaining an unwanted error-marked dataset. Marked data will be stored in the marked dataset and removed from the unmarked dataset. In fact, with the development of the DL course, the dataset gets larger and larger. Meanwhile, more and more data are marked for training DL models.

In our DL course, Dunhuang data is used as a dataset for dataset making. The Dunhuang images include many types, such as People, Animal, Buddha, Cloth, Apsaras, and Building, etc., as shown in Figure 3. We have collected and marked nearly 60 thousand pictures in the DL course. In the future, more types of datasets can be made and shared through the platform.

3.3. Computing resources sharing

For the characteristic of large computational quantity, DL is difficult to be practiced in college education. For instance, training a ResNet50 (He et al., 2015) DL model for image recognition will take about 14 days on Nvidia M40 GPU, and it will take decades to train such a DL model on a regular computer without the help of GPU. Hence, we design a cloud-based computation resources sharing mechanism to provide a high-performance platform. There has been some could-based computation platform for DL model training. For example, Colab and Kaggle are online platforms released by Google, providing computation for machine learning and DL researchers. It is free for ordinary DL users and provides some application examples. However, both Colab and Kaggle limit training resources in a session. Colab offers 12 hours of training timing, and Kaggle delivers 9 hours of training timing, which leads to the use of these two platforms is not flexible enough for some time-consuming training tasks. Another problem is that both Colab and Kaggle are not available for Chinese students due to the national policy.

In our design, we assume that there are 40 students in a class. The shared computation server is configured with Intel Core i9-7900X CPU and four Nvidia GeForce RTX 2080Ti GPUs. Each student is assigned a Jupyter notebook account for DL model training. According to our test, the average performance of each account is stronger than Colab and Kaggle, and students completed their experiments with the help of our shared computation platform. We hope more education institutions join and share the computing resources.

3.4. DL models and transfer learning in DL course

In our DeepFlying platform, we provide DL models sharing mechanism to reuse trained DL models. One common way to solve the computation resource problem is the so-called transfer learning. Transfer learning is often used to reduce the computation based on previous trained DL models when training a DL model. Generally, to train a DL model of similar tasks that exist trained model, we do not need to train the DL model from scratch. The previously trained model can be used to initialize the weights of the DL model. According to the application scenarios, there are four types of transfer learning in DL, as shown in Table 1.

Table 1. Four application scenarios of transfer learning in DL

Sample sizes	Application similarity	Training method
Big	High	Fine tuning
Big	Low	Fine tuning or retraining
Small	High	Modify and train fully-connected layers
Small	Low	Redesign and retrain model

For instance, in the first case, the student can reuse an existing object recognition model by developing an object recognition application whose categories are commonly seen. They only need to download the weights file and load it into the DL model. Then, fine-tuning the DL model based on the new training dataset will achieve a good recognition model. This is the best condition for transfer learning.

In the second case, the new application dataset for training is large, lacking trained similar DL models. This condition mainly appears in specific applications, such as medical image diagnosis. In this case, we can still choose a similar DL model and execute a fine-tuning operation. If the newly trained model has a poor performance, we need to retrain the DL model based on the new dataset.

Another common scenario is that only a small dataset can be obtained, and there are already related DL applications. For instance, we need to develop a new face recognition application used to recognize 30 students. DL based face recognition approaches are widely researched and trained. In this case, the trained DL model of face recognition can be transplanted with modified fully-connected layers. Then, the face recognition model can be trained with the new dataset. In this process, the convolutional layers and pooling layers of trained DL already have feature extraction ability. Hence, the weights of convolutional layers are fixed, and the pooling layers have no weights. Only two or three fully-connected layers on the top need to be trained. It will significantly reduce the amount of computation of model training.

The final scenario of transfer learning is the worst condition, in which there is little training dataset, and there is no similar application model could be referenced. In this case, students need to redesign and train the DL model from scratch. In fact, it seldom occurs from the perspective of education because the dataset and model will accumulate gradually with the launching of the DL course. We can use transfer learning in DL courses and, first, roughly train a DL model with part of the training data before the DL course. Then students can utilize transfer learning and other training data to train and optimize the DL model. That is, we train a DL model as the base model, which is used to make transfer learning by students. The base DL model will be updated with the development of the course. In fact, transfer learning is an important learning approach used by people in the study. It provides a way for students to infer other things from one fact. The applications of transfer learning in DL education can be furtherly researched.

4. DL Course practice

To practice the collective intelligence idea and share learning mechanism for DL education, based on the DeepFlying platform, we start a short-term DL course practice in the summer, which aims to teach students DL technologies, including Convolutional Neural Network (CNN) (Jiang & Chi, 2019), Neural Style Transfer (NSF) (Gatys et al., 2016) and Generative Adversarial Networks (GAN) (Creswell et al., 2017). Based on these technologies, students need to make their artworks in groups.

4.1. Content of DL course

We assume that the students who take the DL course by default have a certain programming foundation, and they know how to program by Python programming language. However, the foundation of DL is unnecessary, and the DL course will be completed in five days.

The history of neural networks and DL will be introduced on the first day. And the disadvantages of neural networks in each stage are presented and discussed. Students are taught that computing power and large amounts of field training data are two critical factors for DL development. The Dunhuang datasets and shared computing resources in the DeepFlying platform are also introduced to students. Finally, students need to design and train a three-layer neural network for simple function fitting based on the TensorFlow DL library with the help of the DeepFlying platform.

The main course content of the second day is about CNN. CNN is the core technology of DL, which utilizes convolutional layers and pooling layers to extract features. Meanwhile, the weight-sharing mechanism of convolutional kernels can effectively reduce the number of weights to reduce the computational complexity of model training. Students need to know how the convolution and pooling operations are conducted today. Also, they need to understand the basic principles to design a DL model. As a simple example, students must implement a handwritten number recognition program based on DL and MNIST dataset (Deng, 2012).

On the third day, NSF and GAN models are introduced to the students. NSF is a type of transfer learning based on the trained DL model. The latest research on deep learning showed that different layers of the DL model could effectively extract features from different levels. Therefore, NSF can combine the features extracted from different images to transfer the style of one image to another. Trained DL models used to extract image features have been provided on the DeepFlying platform. Through NSF model practice, each group can generate their artworks.

On the fourth day, the teaching focal points of the GAN model consist of a Generator and a Discriminator. Students need to know that the Generator and the Discriminator are alternately trained until the Discriminator cannot judge whether the Generator generates the picture or not. Then, the Generator learned the statistical rules of trained images and could develop new artworks. Students can search the varieties of GAN to create their paintings.

4.2. Students attending DL course

We selected 40 students from universities to attend the DL course. These students are from 23 universities, and each university has no more than two students. Therefore, these students are in different educational environments. Four students are randomly selected and assigned into a group, and there are ten groups in total. At the beginning of the course, every group elects a leader by playing a game. These students were arranged to finish the DL course in groups, and each group needed to develop their DL artwork.

4.3. Dataset used in DL course

There is a little difficult when students build their own DL model using open-sourced DL packages, like TensorFlow, Caffe, and Torch. In the meantime, they can find many ready-to-go models on the Internet. However, it's challenging for students to perform DL model when they have limited data for some specific problem. Our course encouraged students to share in this situation, so we designed this dataset marking system as an instance.

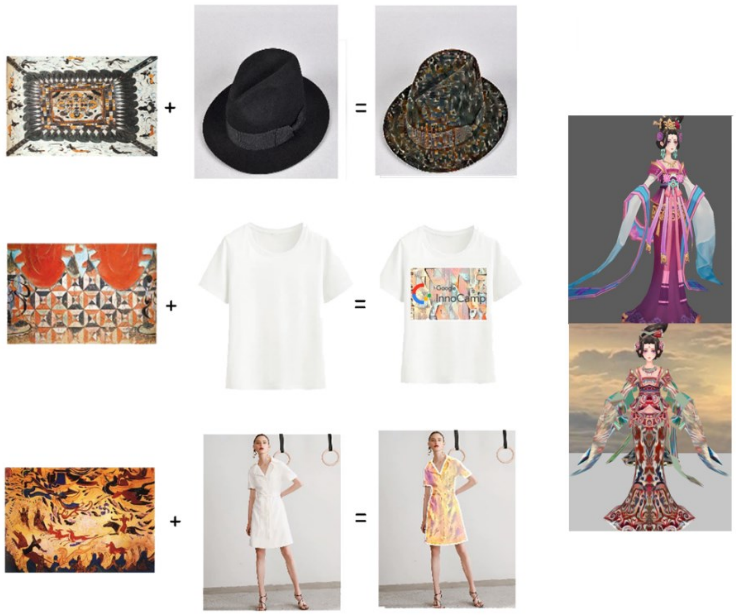
The dataset in the DL course is based on the Dunhuang frescoes images. Dunhuang murals are the representatives of ancient Chinese grotto murals. Although these murals have had a history of thousands of years, they have shown different styles in painting images and color forms. They utilized a small number of colors to paint pictures, which did not make people feel monotonous. In contrast, they show a unique gorgeous. Dunhuang frescoes reflect ancient Chinese culture and art style, which can be well used for today's art design reference. There are many Dunhuang frescoes images online. Students are arranged to crawl these Dunhuang images in groups. The crawled images are messy and need to be selected and sorted out. This is conducted on our dataset system. Students need to mark images online to finish their work. Dunhuang's mural images show different characteristics in the various dynasties. They can also be divided into different categories according to their contents, such as flying apsaras, landscape painting, Buddhistic stories, etc. Our DL course mainly focuses on the characteristic of colors and costumes, which can be used to conduct style transfer and costume generation tasks.

In addition, Dunhuang frescoes images are openly accessible by government-owned websites in China. However, the sites are designed to show art to the masses instead of providing data to DL research, so we crawled and marked the images to make Dunhuang data available to many DL learners.

4.4. Outputs of the DL course

After the theory course, each group is given two days to create their artwork based on NSF and GAN. In fact, students are entirely free to design their artworks. In this process, group leaders assign a task for each student. For example, some students are arranged to collect image materials, while others are engineered to debug or modify the DL code. Finally, each group must submit their artwork and their project document. Some artworks submitted by students are shown in Figure 4.

Figure 4. Some artworks submitted to the DL course



4.5. Evaluation of the artworks

We evaluated these artworks by group voting. Each group chooses a representative to evaluate the artwork of the other groups, and these representatives give scores between 0 and 100. The score results are shown in Table 2. We removed the maximum and minimum scores to get the final average scores of each group, and the top three groups were rewarded. Broke off both ends, the average scores fell in the interval 81 to 86, which shows that most groups learned about similar DL knowledge after the course. The top-ranked groups contain at least one amazing member expert who is good at coding or painting. For the students with lower performance, they need to struggle harder to finish their DL course. Finally, they can complete all courses and create an art piece with the help of their group, even if they have less prior knowledge.

Table 2. Evaluation results of student artworks

Judge/Group	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
J1		77	86	70	90	89	89	83	87	65
J2	88		76	77	75	74	78	75	76	73
J3	100	78		86	88	74	84	84	84	67
J4	98	83	89		87	95	84	80	82	82
J5	99	87	91	92		93	94	92	88	85
J6	94	87	86	92	88		88	87	85	85
J7	95	76	80	82	78	82		80	77	67
J8	92	73	77	78	84	82	85		70	65
J9	90	83	84	79	83	83	85	80		69
J10	95	88	85	89	87	89	86	86	85	
Highest	100	88	91	92	90	95	94	92	88	85
Lowest	88	73	76	77	75	74	78	75	70	65
Total	851	732	754	745	760	761	773	747	734	658
Average	94.71	81.57	83.36	82.29	85.0	84.57	85.86	82.86	82.29	72.57
Rank	1	9	5	7	3	4	2	6	8	10

4.6. Evaluation of the DL course

Two weeks after the DL course, we collected the student opinions by answering the survey online. Finally, 39 questionnaires were obtained (One student did not reply). The main contents of the questionnaire include the evaluation scores of many course items, including registration process evaluation, reception service evaluation, teaching method evaluation, project content evaluation, and accommodations evaluation. The total score of each evaluation is 5. And the statistical results, including total, average scores, and standard deviation of scores, are shown in Table 3.

Table 3. Course evaluation results based on survey online

Score	Total	Average	Std
Registration	188	4.821	0.4456
Reception	192	4.923	0.2664
Teaching method	193	4.948	0.2206
Project content	185	4.743	0.5869
Accommodations	190	4.871	0.3343

In this table, total scores, average scores and the standard deviation of each item are listed. Overall, the evaluation of the DL course is good. And the teaching method got the highest score, which indicates that this teaching method is acceptable for college students, even in a few days.

5. Discussion and conclusions

DL has been the core technology of AI. However, practical DL education needs to solve the course materials, dataset, and computing problems. Therefore, how to promote DL education in college is still a problem due to the complexity of DL course practice. In this article, we proposed the collective intelligence idea for DL materials. Three stages are designed before DL education. Meanwhile, a DeepFlying platform and the resource sharing mechanism to solve data and computation problems are developed. Training datasets are collected and marked by students based on the DeepFlying platform. Some previous trained DL models and computation resources are also provided and shared on the platform.

To test the effect of DL education, we make a short summer school practice. Forty college students are selected to attend the DL course. The course theme mainly focuses on Dunhuang mural pictures. Students are required to collect and classify part of Dunhuang images to put datasets marking into practice, which can also help the platform collect more shared datasets. The contents of the practical DL course mainly include the principles and practice of neural networks, NSF, and GAN. Students submitted their artworks made on the DeepFlying platform according to what they learned in the DL class. Evaluations of the artworks and course are also collected. The results show that students can learn DL contents and finish their artworks well. Meanwhile, they also have a high evaluation for the DL teaching method.

In this paper, we mainly focus on DL education for college students. And the course content is limited to classical CNN-based deep neural network models, which are determined by the background of Dunhuang art. Other deep models, such as long short-term memory (LSTM), can be introduced in the DL course. But it needs more application scenarios and course design. In the future, more applications can be developed for the DL course. In this way, the platform will collect more shared datasets for DL education.

Statements on open data, ethics and conflict of interest

A request to access data can be directed to authors. The research performed in this work is the sole work of the named authors. The ideas presented in this article do not pose any risks to individuals or institutions. We declare that we do not have any conflicts of interest regarding the study.

Acknowledgement

This work was supported by Ministry of Education - China Mobile Research Foundation under Grant No. MCM20170206, The Fundamental Research Funds for the Central Universities under Grant No. lzujbky-2021-sp43, National Natural Science Foundation of China under Grant No. 61402210, State Grid Corporation of China Science and Technology Project under Grant No. SGGSKY00WYJS2000062 and No. SGGSKY00WYJS2100164, Program for New Century Excellent Talents in University under Grant No. NCET-12-0250, Google Research Awards and Google Faculty Award.

References

Ayinde, B. O., Inanc, T., & Zurada, J. M. (2019). Regularizing deep neural networks by enhancing diversity in feature extraction. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(9), 2650-2661. https://doi.org/10.1109/TNNLS.2018.2885972

Chao, Y., Chen, X., & Xiao, N. (2019). Deep learning-based grasp-detection method for a five-fingered industrial robot hand. *IET Computer Vision*, *13*(1), 61-70. https://doi.org/10.1049/iet-cvi.2018.5002

China Daily (CD). (2019). *AI major approved for 35 Chinese universities*. http://www.chinadaily.com.cn/a/201905/24/WS5ce7690da3104842260bd9b3.html

Creswell, A., White, T., Dumoulin, V., Kai, A., Sengupta, B., & Bharath, A. A. (2018). Generative adversarial networks: An Overview. *IEEE Signal Processing Magazine*, *35*(1), 53-65. https://doi.org/10.1109/MSP.2017.2765202

Daniel, B. (2015). Big data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, *46*(5), 904-920. https://doi.org/10.1111/bjet.12230

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Li, F. F. (2009). Imagenet: A Large-scale hierarchical image database. *Proceedings of IEEE Computer Vision & Pattern Recognition*, 248-255. https://doi.org/10.1109/CVPR.2009.5206848

Deng, L. (2012). The MNIST database of handwritten digit images for machine learning research [best of the web]. *IEEE Signal Processing Magazine*, *29*(6), 141-142. https://doi.org/10.1109/MSP.2012.2211477

Freitas, S. I., Morgan, J., & Gibson, D. (2015). Will MOOCs transform learning and teaching in higher education? Engagement and course retention in online learning provision. *British Journal of Educational Technology*, *46*(3), 455-471. https://doi.org/10.1111/bjet.12268

Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2414-2423. https://doi.org/10.1109/CVPR.2016.265

Giannakos, M. N. (2013). Exploring the video-based learning research: A Review of the literature. *British Journal of Educational Technology, 44*(6), E191-E195. https://doi.org/10.1111%2Fbjet.12070

Hameed, A. A., Karlik, B., & Salman, M. S. (2016). Back-propagation algorithm with variable adaptive momentum. *Knowledge-Based Systems, 114,* 79-87. https://doi.org/10.1016/j.knosys.2016.10.001

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 710-778. https://doi.org/10.1109/CVPR.2016.90

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A., & Jaitly, N., Senior, A., Vanhoucke, A., Nguyen, P., Sainath, T. N., & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The Shared views of four research groups. *IEEE Signal Processing Magazine*, *29*(6), 82-97. https://doi.org/10.1109/MSP.2012.2205597

Jiang, Y., & Chi, Z. (2019). A CNN model for semantic person part segmentation with capacity optimization. *IEEE Transactions on Image Processing*, *28*(5), 2465-2478. https://doi.org/10.1109/TIP.2018.2886785

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, *25*, 1097-1105. https://doi.org/10.1145/3065386

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444. https://doi.org/10.1038/nature14539

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324. https://doi.org/10.1109/5.726791

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollr, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. *In European conference on computer vision*, 740-755. https://doi.org/10.1007/978-3-319-10602-1_48

Liyanagunawardena, T. R., Lundqvist, K. Ø., & Williams, S. A. (2015). Who are with us: MOOC learners on a FutureLearn course. *British Journal of Educational Technology*, *46*(3), 557-569. https://doi.org/10.1111/bjet.12261

Lo, J. J., & Shu, P. C. (2005). Identification of learning styles online by observing learners' browsing behaviour through a neural network. *British Journal of Educational Technology*, *36*(1), 43-55. https://doi.org/10.1111/j.1467-8535.2005.00437.x

Pinkus, A. (1999). Approximation theory of the MLP model in neural networks. *Acta Numerica*, *8*, 143-195. https://doi.org/10.1017/S0962492900002919

Terras, M. M., & Ramsay, J. (2015). Massive open online courses (MOOCS): Insights and challenges from a psychological perspective. *British Journal of Educational Technology*, *46*(3), 472–487. https://doi.org/10.1111/bjet.12274

Tingting, L., & Mengyu, X. (2019). Analysis and evaluation on the quality of news text machine translation based on neural network. *Multimedia Tools and Applications*, *79*(23), 17015–17026. https://doi.org/10.1007/s11042-019-7532-5

Welham, D. (2008). AI in training (1980–2000): Foundation for the future or misplaced optimism? *British Journal of Educational Technology*, *39*(2), 287-296. https://doi.org/10.1111/j.1467-8535.2008.00818.x

Ye, Y. Y., Hao, X. L., & Chen, H. J. (2018). Lane detection method based on lane structural analysis and CNNS. *IET Intelligent Transport Systems*, *12*(6), 513-520. https://doi.org/10.1049/iet-its.2017.0143

Zhi, J., Sun, J., Wang, Z., & Ding, W. (2018). Support vector machine classifier for prediction of the metastasis of colorectal cancer. *International Journal of Molecular Medicine*, *41*(3), 1419-1426. https://doi.org/10.3892/ijmm.2018.3359

Lin, J., Sun, G., Beydoun, G., & Li, L. (2022). Applying Machine Translation and Language Modelling Strategies for the Recommendation Task of Micro Learning Service. *Educational Technology & Society*, *25* (1), 205-212.

Applying Machine Translation and Language Modelling Strategies for the Recommendation Task of Micro Learning Service

Jiayin Lin[1*], Geng Sun[1], Ghassan Beydoun[2] and Li Li[3]

[1]University of Wollongong, Wollongong, NSW, Australia // [2]University of Technology Sydney, Sydney, NSW, Australia // [3]Southwest University, Chongqing, China // jl461@uowmail.edu.au // gsun@uow.edu.au // ghassan.beydoun@uts.edu.au // lily@swu.edu.cn

[*]Corresponding author

ABSTRACT: A newly emerged micro learning service offers a flexible formal, informal, or non-formal online learning opportunity to worldwide users with different backgrounds in real-time. With the assist of big data technology and cloud computing service, online learners can access tremendous fine-grained learning resources through micro learning service. However, big data also causes serious information overload during online learning activities. Hence, an intelligent recommender system is required to filter out not-suitable learning resources and pick the one that matches the learner’s learning requirement and academic background. From the perspective of natural language processing (NLP), this study proposed a novel recommender system that utilises machine translation and language modelling. The proposed model aims to overcome the defects of conventional recommender systems and further enhance distinguish ability of the recommender system for different learning resources.

Keywords: Information filtering, Recommender system, Micro learning, Big data, Natural language processing

1. Introduction

The achievements in 5G Internet and mobile devices boost the real-time multi-media interaction in various applications such as commercial, entertainment, and online education. In the meantime, the fast-paced modern life and booming of knowledge in the big data era drive people to seek a more flexible way to acquire knowledge or carry out personalised learning activities. All the above factors gave birth to micro learning service (Sun et al., 2015b), aiming to utilize user’s daily fragmented spare time and assist the learner in conducting self-regulated personalized learning activities. The term “micro” used in this study refers to the small (micro) personalised chunks of learning materials containing a small volume of knowledge. As pointed out in the study (Syeda-Mahmood & Ponceleon, 2001), users are less likely to leave out the knowledge points for a short learning session, such as a short video. And the engagement of an online learning activity plunges quickly after 7 minutes (Anderson et al., 2014; Guo et al., 2014). With the advantage of Internet technology, massive learning materials are uploaded to the Internet every day in various disciplines, format, and difficulty levels. Hence, a serious information overload problem challenges the learner experience of the micro learning service. Hence, filtering out irrelevant information and picking the one that matches the learner’s learning requirement is the key to such a personalized online learning service.

Especially for the online service/application that deploys in the context of big data, a sophisticated recommender system is a key factor to guarantee efficiency and personalization. Even information filtering and retrieval were classified into two different research disciplines, the boundary between information filtering (i.e., the main function of recommender system) and information retrieval (i.e., the main function of search engine) is relatively vague. The former one aims to find irrelevant resources and filter them out, and the recommenders can be further classified into three categories (Wasid & Ali, 2017): content-based filtering (CB), collaborative filtering (CF), and hybrid recommending strategy. The latter one aims to find relevant resources and rank them based on their relevant degrees. In General, both of them try to distinguish relevant and irrelevant information (Belkin & Croft, 1992). Due to pedagogical issues (Lin et al., 2020), ranking the recommended learning materials is significant to delivering a suitable learning resource to the learner. However, as discussed in (Valcarce, 2015), there is little research about applying information retrieval (IR) ideas to boost the performance of a recommender system (RS). From the perspective of IR, the probabilistic model has a solid statistical foundation. Hence, as discussed in (Belkin & Croft, 1992) that the probabilistic model can make significant improvements to the field of recommender system as it did in IR.

Based on one previous work (Lin, 2020), in this paper, we further refine the proposed recommender system, which can precisely rank the recommended learning materials. This system makes use of the mathematical concepts of machine translation and language modelling to model the learning materials and reflect the mapping relationships between historical learning records and new learning materials. The remainder of the paper is organised as follows. Section 2 will discuss challenges in the recommendation task of the micro learning service.

The related work in this research will be discussed in Section 3. The proposed model will be introduced in Section 4. We introduce and explain the evaluation of the proposed recommendation strategy in Section 5. The conclusions and future work of this study is discussed in Section 5.

2. Challenges in recommendation for micro learning

2.1. The drawbacks of conventional recommendation models

As discussed before, the algebraic-based recommender system such as collaborative filtering and matrix factorization has been demonstrated to be very effective in filtering out irrelevant online resources, but such a model lacks the ability to precisely distinguish the difference between the remained resources. Most of the system used algebraic-based strategy can only predict the rating value of the resources but cannot provide any detail information of the resource with the same rating value. One study proved that the algebraic-based collaborative filtering cannot generally provide good result in the top-k recommendation task (Valcarce, 2015). Hence, the authors of this study argued that the probabilistic method could be a more effective and formal way for generating personalized rankings of recommendations (Valcarce, 2015). And in the study of (Koren et al., 2009), the researchers demonstrated that matrix factorization based recommender systems is guided by the rating value and does not involve any explicit features, which could not represent the ranking information among items.

2.2. Micro learning service and recommendation in e-learning

Most studies on recommenders found in the e-learning field were focused on the suitability of learning materials against learners' personalization. Formally, a micro learning activity is carried out within a time span of 15 minutes through a mobile device (typically, though). One pilot work investigated the possibility of customising open educational resources (OERs) to meet the demand of microlearning (Sun et al., 2015b). And another work provided a comprehensive learner model oriented to micro learning through OERs in (Sun et al., 2015a). In (Sun et al., 2018) the researchers discussed the mainstream typology (video, audio, text), type of interaction (expositive, active, mixed, two-way), didactic model (e.g., inductive, deductive, learning by doing) of the online learning materials, in particular, for micro learning.

A content-based convolutional neural network (CBCNN) recommender system was proposed in a prior study (Shu et al., 2018), which shows fairly satisfying ability in mining new or unpopular learning materials for a target learner. Another study proposed a new way to calculate similarities between online learning materials for recommendation tasks (Niemann & Wolpers, 2013). And the authors in that paper argued that the usage context-based model has the potential to outperform the content-based model, if the usage data is sufficiently fine-grained. And a system for recommending OERs in MOOC was proposed in (Hajri et al., 2017), which emphasized the significance of modelling users and learning materials. However, none of these studies mentioned the significance of the ranking for the success of an online learning service.

2.3. The significance of ranking ability of the recommender system for online learning

Unlike the personalized service in other areas (e.g., e-commerce and entertainment), complex pedagogical issues (Sikka et al., 2012; Wu et al., 2015) or requirements influence the learning outcome to a great extent. For example, the description of a learning material might contain vague information and pre-requested knowledge are required for some courses. Letting learner know what he/she should learn first what he/she needs to learn next is vital for an informal or non-formal online learning. Hence, for the online learning service like micro learning, a recommender system should be able to precisely distinguish the importance differences of the recommended resources (ranking).

3. Related work

3.1. Conventional recommendation strategies

Collaborative filtering and content-based filtering are two typical conventional recommendations strategies, which have been proven as effective and been commonly used in many studies or real applications. The recommendation results for a target user given by collaborative filtering are based on his/her correlation among other users of the system. As indicated in one previous study (Pazzani, 1999), collaborative filtering presents a uniform approach to finding items of potential interest and predicting the rating that the target user would give to the item.

Content-based filtering strategy generates recommendations by comparing and analysing the description of the items that have been rated by the target user and the descriptions of the items to be recommended (Pazzani, 1999). However, as the user's profile is constructed based on the user's historical activities, such recommendation strategy lacks the ability to explore and recommend the new items, which might vary greatly from the historical ones.

3.2. Language model and translation model for information retrieval and information filtering

As discussed in Berger and Lafferty (2017), applying the strategy of machine translation to solve the recommendation problem is not a fanciful idea but a feasible one. In this study, the researchers demonstrated constructing using a statistical machine translation model to handle the information retrieval task. Similarly, in another research (Lavrenko & Croft, 2017), researchers used a language model to reflect the mapping relationships between a query and a document. Statistical language models were explored and analysed for handling the recommendation task (Valcarce, 2015). However, applying a language model a translation model to solve a recommendation problem is still less touched.

4. Computation model

In this section, we roll out a novel recommendation strategy based on the combination of the concept of language and translation model. This strategy is realised in the recommender module of the proposed system in one early work (Lin et al., 2019a).

4.1. Translation model and language model

4.1.1. Statistical machine translation model

The translation is a probabilistic mapping procedure that a string e in one language can be translated to a string f in another language with the probability of $P(f|e)$. In the natural language processing (NLP) area the probability distribution of $P(f|e)$ can be modelled in different ways. For example, Bayes Theorem is used in one previous study to represent this distribution (Brown et al., 1992):

$$P(f|e) = \frac{P(e)P(f|e)}{P(f)} \qquad (1)$$

Since the denominator only correlates with source language f and we only consider the result of target language e, we can simplify this distribution as Equation (2):

$$P(e|f) \propto P(e)P(f|e) \qquad (2)$$

Finding the best translation result $\hat{e}$ is realised by finding the one that gives the highest probability:

$$\hat{e} = \underset{e}{argmax} P(e)P(f|e) \qquad (3)$$

4.1.2. Language model

Generally, the expression of a language is composed of sentences and phrases, and the representation of sentences or phrases is a sequence of words. The language model is a probability distribution of a sequence of words. In Brown et al. (1992), the authors assumed that the production of a piece of English text could be characterized by a set of conditional probabilities. Given an English sentence or phrase e which contains k words, its probability can be formulated as Equation (4):

$$P(e) = P(w_1, \dots, w_k) = P(w_1)P(w_1|w_2)P(w_3|w_2,w_1) \dots P(w_k|w_{k-1},w_{k-2},\dots,w_1) \qquad (4)$$

N-gram model is one of the most representative language models applied in many NLP tasks, such as speech recognition, spelling correction, and translation. Given a sequence of *n-1* words, the n-gram model predicts the probability of the next word after this sequence. As the n-gram model keeps the continuous combination of n words, it is capable to preserve and represent some semantic information. When using the n-gram model, the probability of producing a sequence of words can be formulated as Equation (5):

$$P(w_1, \dots, w_k) = P(w_1)P(w_1|w_2)\dots P(w_k|w_{k-1}, w_{k-2}, \dots, w_1) \approx \prod_{i=1}^{k} P(w_i|w_{i-(n-1)}, \dots, w_{i-1}) \quad (5)$$

The probabilities used in these models can be simply calculated by using maximum likelihood estimation (MLE), for n-gram models and the translation procedures the probability can be formulated as Equation (6):

$$P\left(w_i|w_{i-(n-1)}, \dots, w_{i-1}\right) = \frac{C(w_{i-(n-1)}, \dots, w_i)}{C(w_{i-(n-1)}, \dots, w_{i-1})} \quad (6)$$

Here, $C(w_1, \dots, w_i)$ represents the frequency of the word sequence $w_1, \dots, w_i$ in the training sample.

4.2. The combination of the language model and the translation model

As discussed in the early pilot study (Lin, 2020), the language model can be used to model the online learning materials and historical learning records, and the machine translation model can be used to model the mapping relationship between historical learning records and the new learning materials. More specifically, in NLP, a language model is used to reflect the combination between words, while the proposed system utilizes a language model to reflect the combination between features of a certain learning material. For a translation task, a machine translation model is used to reflect the mapping relationship between two different languages, while the proposed system uses a machine translation model to reflect the mapping relationship between historical learning records and new learning materials.

The visualization of the process of using a language model to reflect the combination between words and feature combination of a learning material is shown in Figure 1. We can see that a sentence is composed of a sequence of words ($w_1, w_2, w_3, \dots, w_n$), and similarly, a learning material is represented by several features ($f_1, f_2, f_3, \dots, f_n$) such as type, language, discipline etc. For the recommendation process that tackled in this study, it is represented through a mapping procedure between the historical learning records and the new learning material, such procedure is similar to the process of language translation. The visualization of the translation procedure is shown in Figure 2. In the left part of Figure 2 is the translation procedure between one sentence in the original language and a set of translation results in the target language. Different translation results have different probabilities. The result with a higher probability means the result is the more reasonable one Similarly, in our proposed solution, the recommendation procedure between the historical learning activities and the recommended new learning materials is represented in the right part of Figure 2. Given the historical learning activities of a user, the system will recommend several different new learning materials with different probabilities. The higher probability of the learning material means it is more suitable to this user.

Figure 1. The modelling process of a sentence and a piece of learning material

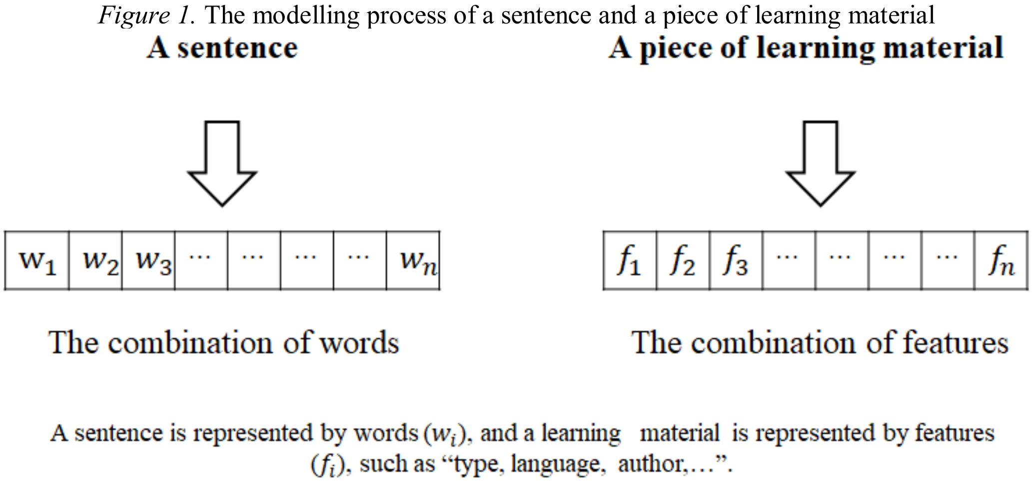
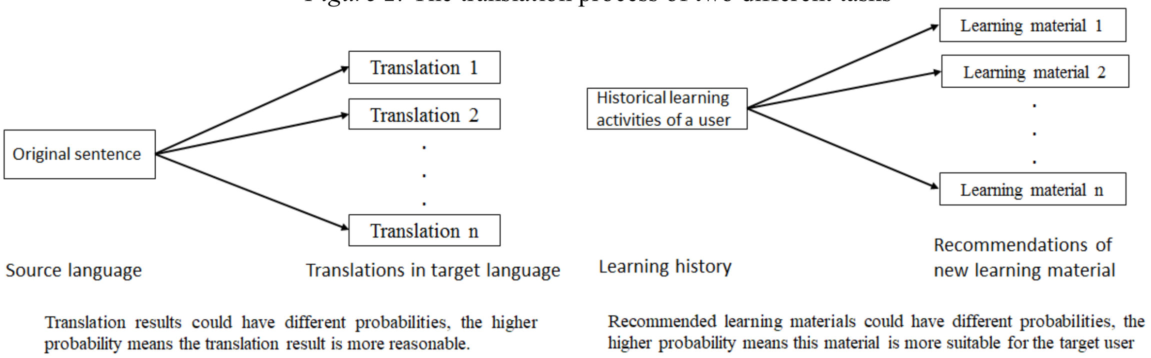


Figure 2. The translation process of two different tasks



Hence, the proposed recommender system in this study is composed of a language model and a machine translation model. The recommendation process is formulated as Equation (7) below:

$$P(l|h) \propto \prod_{i=1}^{k} P(f_i|f_{i-(n-1)}, \dots, f_{i-1}) P(h|l) \qquad (7)$$

Herein based on the concepts defined in Lin (2020), l represents the learning material, h represents the historical learning records of the target user, f_i is the i-th feature that is used to represent the leaning material. The probability $P(l|h)$ represents the degree of correlation between a user's historical learning activities and the new learning material. Finding the best recommendation result $\hat{l}$ is realised by finding the one that gives the highest probability which is formulated as Equation (8):

$$\hat{l} = \underset{l}{argmax} \prod_{i=1}^{k} P\left(f_i \middle| f_{i-(n-1)}, \dots, f_{i-1}\right) P(h|l) \quad (8)$$

4.3. Sub-translation for different types of features

As the different types of features contain different amount of information, it is more reasonable to interpret different types of feature/metadata separately. For example, some descriptive features, such as a subject title and the introduction of a course, are more important than the metadata (e.g., the resolution degree of the lecture video). However, such less-descriptive metadata can also reflect some latent information of the target user (Al-Hmouz et al., 2011), to some extent. During the recommendation process, different types of feature/metadata should be treated in different manners. Hence, it is more reasonable that the translation procedure proposed above is separated into several sub translation tasks, then all the translation results are assembled with different weighting values. For linearly assembling the several translation results, the final relevance score s between the recommended learning material and the user's historical learning records can be estimated by the following equation:

$$s = \sum_i \alpha_i p_i \text{ , and } \sum_i \alpha_i = 1 \qquad (9)$$

Herein, α_i are the weight for the i-th translation result. As described in Equation (7), p_i represents the correlation degree between the recommended learning material and the user's historical learning records produced by the i-th translation task. The top k recommended items are generated by ranking the s scores and picking the k items with the highest values. The workflow of our proposed translation set is shown in Figure 3.

5. System evaluation

In this section, we will discuss the feasibility analysis of our proposed model, the dataset used in the experiments, and the relevant evaluation metrics.

5.1. Feasibility analysis

The analysis of the feasibility of the proposed model and relevant experiments stems from three perspectives, model design, experimental dataset, and the evaluation metrics.

The proposed recommendation strategy does not involve any less-explainable “black-box” structure, such as the neural network. All the model designs are based on the Naïve Bayes rule and its variants, and all the formula deductions discussed in the previous sections are based on the solid mathematical foundation and probability theory.

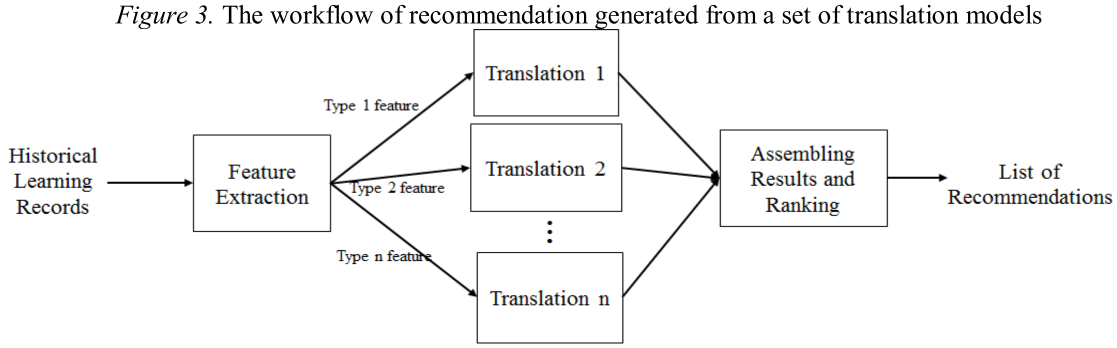
The dataset used in the relevant experiments of this study is a well-acknowledged public dataset, which is widely used in various recommender system related studies. The details of such dataset will be discussed in the next subsection.

In our work, we do not involve any novel evaluation metrics, all the evaluation metrics used in the experiments are also well-acknowledged in the research area of the recommender system.

5.2. Dataset

In one earlier work, researchers investigated the readiness of public and academic data sources that were adopted in e-learning literature (Lin et al., 2019b). By comparing the pros and cons of available public data sets, the dataset used in the experiments of this study is Book-Crossing (Ziegler et al., 2005). For the experiment, we also crawled some other descriptive information like description and comments for each book. The reasons of choosing this dataset can be summarized as follow:

Figure 3. The workflow of recommendation generated from a set of translation models



Compared to other open-source datasets like MovieLens (Harper & Konstan, 2015) and Jester (Goldberg et al., 2001), Book-Crossing dataset is more closer to the educational domain.

This work aims to demonstrate and prove that the translation and language models have the potential in mining latent useful information for boosting the recommending results. To demonstrate the effectiveness of applying this recommending strategy to the different formats of online learning material is beyond the scope of this paper.

5.3. Evaluation metrics

5.3.1. Normalized Discounted Cumulative Gain (NDCG)

As the proposed system aims to precisely rank the recommended learning materials, comparing the proposed system’s ranking ability with the baseline is necessary. Hence, Normalized Discounted Cumulative Gain (NDCG) (Wang et al., 2013) will be used for measuring the ranking performance of the baselines and the proposed system.

5.3.2. Precision and recall at top K

Moreover, for a comprehensive comparison, we will also use Precision@K and Recall@K to evaluate the models for top K recommended results. These two metrics reflect the ability of a model to find relevant learning resources from the online repository.

6. Conclusions and future work

In this paper, we proposed a novel recommender system which applies the idea of language modelling and machine translation. From mathematic derivation, we can see that the proposed system can distinguish the importance differences of the recommended results. In the future, we will conduct the experiment of the proposed model based on the dataset from the educational domain. In this study, for assembling the set of translators, we proposed a simple linear assembling strategy. However, as demonstrated in one relevant study (Sagi & Rokach, 2018), some non-linear ensemble strategies also have strong ability to discover the complex structure and learn high-level concepts in large datasets. Hence, investigating how to better integrate sub-translators effectively is another area for our future work.

References

Al-Hmouz, A., Shen, J., Al-Hmouz, R., & Yan, J. (2011). Modeling and simulation of an adaptive neuro-fuzzy inference system (ANFIS) for mobile learning. *IEEE Transactions on Learning Technologies, 5*(3), 226-237.

Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2014). Engaging with massive online courses. In *23rd International Conference on World wide Web* (pp. 687–698). https://doi.org/10.1145/2566486.2568042

Belkin, N. J., & Croft, W. B. (1992). Information filtering and information retrieval: Two sides of the same coin? *Communications of the ACM, 35*(12), 29-38. https://doi.org/10.1145/138859.138861

Berger, A., & Lafferty, J. (2017). Information retrieval as statistical translation. *ACM SIGIR Forum, 51*(2), 219–226. https://doi.org/10.1145/3130348.3130371

Brown, P. F., Della Pietra, V. J., Desouza, P. V., Lai, J. C., & Mercer, R. L. (1992). Class-based n-gram models of natural language. *Computational Linguistics, 18*(4), 467-480.

Goldberg, K., Roeder, T., Gupta, D., & Perkins, C. (2001). Eigentaste: A Constant time collaborative filtering algorithm. *Information Retrieval, 4*(2), 133-151.

Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An Empirical study of MOOC videos. In *Proceedings of the first ACM conference on Learning@ scale conference* (pp. 41–50). https://doi.org/10.1145/2556325.2566239

Hajri, H., Bourda, Y., & Popineau, F. (2017). MORS: A System for recommending OERs in a MOOC. In *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)* (pp. 50-52). IEEE. https://doi.org/10.1109/ICALT.2017.89

Harper, F. M., & Konstan, J. A. (2015). The movielens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems (TIIS), 5*(4), 1-19. https://doi.org/10.1145/2827872

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer, 42*(8), 30-37. http://dx.doi.org/10.1109/MC.2009.263

Lavrenko, V., & Croft, W. B. (2017). Relevance-based language models. *ACM SIGIR Forum, 51*(2), 260–267. https://doi.org/10.1145/3130348.3130376

Lin, J. (2020). Hybrid translation and language model for micro learning material recommendation. In *2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)* (pp. 384-386). IEEE. https://doi.org/10.1109/ICALT49669.2020.00121

Lin, J., Sun, G., Cui, T., Shen, J., Xu, D., Beydoun, G., Yu, P., Pritchard, D., Li, L., & Chen, S. (2020). From ideal to reality: Segmentation, annotation, and recommendation, the vital trajectory of intelligent micro learning. *World Wide Web, 23*(3), 1747-1767. http://dx.doi.org/10.1007/s11280-019-00730-9

Lin, J., Sun, G., Shen, J., Cui, T., Yu, P., Xu, D., & Li, L. (2019a). A Survey of segmentation, annotation, and recommendation techniques in micro learning for next generation of OER. In *2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD)* (pp. 152-157). IEEE. https://doi.org/10.1109/CSCWD.2019.8791891

Lin, J., Sun, G., Shen, J., Cui, T., Yu, P., Xu, D., Li, L., & Beydoun, G. (2019b). Towards the readiness of learning analytics data for micro learning. In *International Conference on Services Computing, SCC 2019. SCC 2019. Lecture Notes in Computer Science* (Vol 11515, pp. 66-76). Springer. https://doi.org/10.1007/978-3-030-23554-3_5

Niemann, K., & Wolpers, M. (2013). Usage context-boosted filtering for recommender systems in TEL. In *European Conference on Technology Enhanced Learning, Scaling up Learning for Sustained Impact. EC-TEL 2013. Lecture Notes in Computer Science* (Vol. 8095, pp. 246-259). Springer. https://doi.org/10.1007/978-3-642-40814-4_20

Pazzani, M. J. (1999). A Framework for collaborative, content-based and demographic filtering. *Artificial Intelligence Review, 13*(5), 393-408.

Sagi, O., & Rokach, L. (2018). Ensemble learning: A Survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8*(4), e1249. https://doi.org/10.1002/widm.1249

Shu, J., Shen, X., Liu, H., Yi, B., & Zhang, Z. (2018). A Content-based recommendation algorithm for learning resources. *Multimedia Systems, 24*(2), 163-173. http://dx.doi.org/10.1007/s00530-017-0539-8

Sikka, R., Dhankhar, A., & Rana, C. (2012). A Survey paper on e-learning recommender system. *International Journal of Computer Applications, 47*(9), 27-30.

Sun, G., Cui, T., Guo, W., Beydoun, G., Xu, D., & Shen, J. (2015a). Micro learning adaptation in MOOC: A Software as a service and a personalized learner model. In *International conference on web-based learning, Lecture Notes in Computer Science* (Vol. 9412, pp. 174-184). Springer. https://doi.org/10.1007/978-3-319-25515-6_16

Sun, G., Cui, T., Xu, D., Shen, J., & Chen, S. (2018). A Heuristic approach for new-item cold start problem in recommendation of micro open education resources. In *International conference on intelligent tutoring systems 2018. Lecture Notes in Computer Science* (Vol. 10858, pp. 212-222). Springer. https://doi.org/10.1007/978-3-319-91464-0_21

Sun, G., Cui, T., Yong, J., Shen, J., & Chen, S. (2015b). MLaaS: A Cloud-based system for delivering adaptive micro learning in mobile MOOC learning. *IEEE Transactions on Services Computing, 11*(2), 292-305.

Syeda-Mahmood, T., & Ponceleon, D. (2001). Learning video browsing behavior and its application in the generation of video previews. In *Proceedings of the ninth ACM international conference on Multimedia* (pp. 119–128). https://doi.org/10.1145/500141.500161

Valcarce, D. (2015). Exploring statistical language models for recommender systems. In *Proceedings of the 9th ACM Conference on Recommender Systems* (pp. 375–378). https://doi.org/10.1145/2792838.2796547

Wang, Y., Wang, L., Li, Y., He, D., Chen, W., & Liu, T.-Y. (2013). A Theoretical analysis of NDCG ranking measures. In *Proceedings of the 26th annual conference on learning theory* (COLT 2013), (Vol. 8, pp. 6).

Wasid, M., & Ali, R. (2017). Use of soft computing techniques for recommender systems: An Overview. In *Applications of soft computing for the web* (pp. 61-80). Springer. https://doi.org/10.1007/978-981-10-7098-3_5

Wu, D., Lu, J., & Zhang, G. (2015). A Fuzzy tree matching-based personalized e-learning recommender system. *IEEE transactions on fuzzy systems, 23*(6), 2412-2426. http://dx.doi.org/10.1109/TFUZZ.2015.2426201

Ziegler, C.-N., McNee, S. M., Konstan, J. A., & Lausen, G. (2005). Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web* (pp. 22–32). https://doi.org/10.1145/1060745.1060754

Zhai, X., Chu, X., Meng, N., Wang, M., Spector, M., Tsai, C.-C., & Liu, H. (2022). The Effect of Multi-mode Stimuli of Feedforward and Eye Tracking on Metacognition— An Exploratory Study Using Digital Dictionaries. *Educational Technology & Society*, *25* (1), 213-227.

The Effect of Multi-mode Stimuli of Feedforward and Eye Tracking on Metacognition— An Exploratory Study Using Digital Dictionaries

Xuesong Zhai[1], Xiaoyan Chu[1], Nanxi Meng[2], Minjuan Wang[3], Michael Spector[4], Chin-Chung Tsai[5] and Hui Liu[1*]

[1]College of Education, Zhejiang University, Hangzhou, Mainland China // [2]World languages, Literatures and Cultures, University of North Texas, Denton, US // [3]Learning Design and Technology, San Diego State University, San Diego, US // [4]College of Information, University of North Texas, Denton, US // [5]Program of Learning Sciences and Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, Taipei, Taiwan // xszhai@zju.edu.cn // xiaoyan_chu@zju.edu.cn // nanxi.meng@unt.edu // mwang@sdsu.edu // Mike.Spector@unt.edu // tsaicc@ntnu.edu.tw // autumnhuihui@zju.edu.cn

[*]Corresponding author

ABSTRACT: Metacognition is regarded as a retrospective skill promoting learners' learning performance, deep thinking, and academic well-being. Stimulated Recall (SR) is regarded as a reliable approach to inspiring learners' metacognition in the classroom. However, the outbreak of COVID-19, causing widespread class suspension, may impair the effect of SR on cultivating distance learners' metacognition. The current study, employing multi-mode stimuli of learners' eye movements and feedforward, aimed to develop the effect of SR on activating learners' metacognition in remote settings. Forty-eight university graduates were recruited to participate in an eye-tracking experiment using digital dictionaries. Their feedforward and eye movements were collected as multi-mode stimuli. By reviewing the consistency and discrepancies between their feedforward and eye movements, participants were invited to conduct an SR interview, which stimulated them to retrospect on their prior cognitive behaviors. The results of the metacognition scale pre-post test showed that learners' metacognitive skills were significantly improved by the stimulated recall with multi-mode stimuli. The findings theoretically enrich the metacognition strategy in the Cognitive Theories of Multimedia Learning, and practically extend the implementation of stimulated recall in distance learning contexts.

Keywords: Metacognition, Multi-mode stimuli, Stimulated recall, Eye tracking, Digital dictionary

1. Introduction

The Cognitive Theories of Multimedia Learning (CTML) emphasize the importance of metacognition for multimedia learning outcomes (Moreno & Mayer, 2007), academic well-being (Nasirzade & Nargesian, 2019), and higher-order thinking skills (Parlan & Rahayu, 2021). However, the outbreak of COVID-19 has challenged the cultivation of students' metacognition (Chakma et al., 2021). During the widespread suspension of physical classes, although the rich multimedia materials are options supporting students to conduct remote learning, the overwhelming abundance of the learning materials may distract them from their prior learning goals (Zhang & Zou, 2021). In this situation, students may thus fail to efficiently retrospect prior learning purposes and behaviors, and their metacognition may be simultaneously impaired. In addition, teachers were generally impelled to expend extra efforts to adjust to various learning techniques working online in the pandemic period, so that the retrospective strategy of inspiring students' metacognition has been less investigated (Abdullah, 2020). The above challenges lead to the need for urgent solutions to the problem of inspiring students' metacognition with a reliable approach in the widespread remote learning context.

Metacognition occurs in the condition in which learners make critical judgements on their previous learning behaviors and cognition, by which a meaning retrospection is generated (Taub & Azevedo, 2019). However, a big challenge of cultivating learners' metacognition is that students are used to recollecting the knowledge they have learned rather than retrospecting their prior behaviors and cognition (Rivers, 2020). Stimulated Recall (SR) is regarded as a reliable approach to guiding learners to implement effective retrospection of their prior behaviors, and the selection of the adaptive stimuli is a crucial factor in the successful occurrence of metacognition (Mudrick et al., 2019). Eye-tracking technology has been explored to capture online learners' behaviors as stimuli, by which some specific surveys were conducted to understand online learners' behavior (Wang et al., 2019). Regrettably, this approach may fail to stimulate learners' metacognition, since students answered the survey questions based on what they had already been told about the eye movements, so they recollected the learning process according to their eye movements without retrospection (Horská et al., 2020).

Moreno and Mayer (2007) CTML and Dunlosky's (2005) levels-of-disruption hypothesis suggested that the monitoring of feedforward and disruptions in multimedia are influential cues for metacognition. Constructing a learning setting in which students are able to compare and contrast learning materials and strategies could benefit their metacognition (Rollwage et al., 2018). Feedforward enables the prediction of learning behavior from the future concerning the desired behavior which the subject is encouraged to adopt. The current study proposed to employ stimulated recall with multi-mode stimuli including participants' feedforward and eye movements in an experiment using digital dictionaries. Students were surveyed about their feedforward first, and then their eye-media interactions were captured by eye trackers. In the third stage, by reviewing the consistency and discrepancies of the captured eye movements and learners' own prior feedforward, participants were stimulated to retrospect on what they expected when using digital dictionaries and what their real behaviors were. To examine the effectiveness of the multi-mode stimuli on improving learners' metacognitive skills, a pre- and post-test of metacognition were conducted.

2. Research background

2.1. Metacognition in multimedia learning

According to the Cognitive Theories of Multimedia Learning, a well-designed digital learning environment can significantly improve learning outcomes and perceptions when aligned with learners' cognitive processes, including essential processing, extraneous processing, and generative processing (Mayer, 2014). Essential processing is the first stage where learners get preliminary notification and classification of the presented materials. Then, in the stage of extraneous processing, learners reorganize the current orders, forms, and layout of materials according to their individualized cognitive architecture. Finally, to achieve generative processing, learners need to connect the reorganized material to their feedforward, where their metacognitive skills are aroused. Associated with the neural mechanism, metacognitive function could be examined from the frontal cortex (Frith, 2012), while surveys are regarded as a feasible measurement of metacognition in educational research (Antonietti et al., 2015).

Metacognition, inspiring learners to be aware of their cognitive process, is a higher level of thinking capacity; it is also referred to as "the thinking about thinking" (Renkl et al., 2013). Dunlosky and Metcalfe (2008) defined metacognition as a mental activity of understanding and regulating the learning process, including learners' beliefs about learning, monitoring the state of their knowledge, and controlling their learning activities. CTML emphasized that metacognition significantly impacts problem-solving, reasoning, and academic success in multimedia learning contexts (Mayer, 2014). Some researchers have proposed that metacognitive skills can help learners regulate their learning in online contexts because of their awareness of cognitive processes, and the results found that the more metacognitive skills the learners possessed, the more knowledge could be investigated from multimedia presentation to meet their needs (Antonietti et al., 2015). However, while learning during the COVID-19 outbreak, students confront less supervised environments; it is thus an urgent requirement to inspire learners to use metacognitive mentoring and conduct metacognitive control.

According to the CTML, metacognition may occur when learners connect the multimedia to their prior knowledge in generative processing (Edwards, 2010). CTML encourages comparing and contrasting students' prior knowledge and the current learning behavior in the stage of generative processing in multimedia learning (Mayer, 2014). For example, some researchers have proposed that learners' metacognitive skills were promoted when their retrospection of initial learning goals was awakened (Meuwese et al., 2014). Even though learners' actual behavior may not be consistent with their prior cognition, it also benefits the cultivation of metacognitive skills through their retrospection (Dulamă & Ilovan, 2016). The possible explanation may be due to Dunlosky's (2005) levels-of-disruption hypothesis, which states that the discrepancies within the multimedia content would stimulate learners' comprehension and metacognition when they monitor disruptions and conflicts. Some researchers have employed conflict questions to explore learners' feedback on the understanding of prime numbers, and the results showed that the learners experiencing conflict generated more metacognitive abilities (Questienne et al., 2018). Moreover, the effectiveness of monitoring disruptions relies on the relationship of the multimedia which are displayed, such as the coherence of the verbal and pictorial presentations (Mayer, 2014). Thus, we proposed that learners' feedforward on the functions of digital dictionaries and their eye movements in these areas could be employed as multi-mode stimuli.

2.2. Eye-tracking and multi-mode stimuli

The eye-tracking technique, widely employed in digital learning, is able to capture users' eye movements when they interact with learning materials, by which learners' cognitive behaviors could be observed, examined, and explained (Zhai et al., 2018b). Based on the eye-mind theory hypothesis, the eye-tracking approach allows a dynamic trace of attention to be observed via eye movements (Cortina et al., 2015). Eye movements consist of three basic evaluative criteria: fixation counts, fixation duration, and scanning paths (Lai et al., 2013; Luo et al., 2017). Firstly, fixation count was defined as the concentrations counted in certain Area of Interests (AOIs). According to Rayner (2009), a fixation count lasts over 200 milliseconds. Fixation counts could be seen as a reliable tool to gauge the level of complexity, importance, and viewing. Secondly, fixation duration was defined as the sum of duration of eye movement within certain AOIs that is examined on the time scale. Some researchers pointed out that varied learning motivations and the materials' complexity may influence learners' fixation duration (Park et al., 2015). Typically, the integration of the fixation counts and duration are employed to reflect students' focusing on certain AOIs in the media. Thirdly, the scanning path, presenting the fixations' orders, reveals the holistic logical connection of components, which is adopted to gain access to visual memories in space (Lorigo et al., 2008). Additionally, the developed visualizing technique has facilitated presentation of the eye movements, and the heatmap and scanning figures were typically illustrated to explain learners' perceptual and cognitive process (Wang et al., 2016).

Although the eye-tracking technique is regarded as an adaptable approach to obtaining objective data, it has also been suggested to integrate it with qualitative methods to investigate the driving mechanisms of cognitive processes. Previous studies have found that eye-tracking alone may lead to biased results, and the combination of eye-tracking techniques and a survey could provide a comprehensive understanding of human behaviors (Leszkowicz, 2011). Eye-tracking only tells how learners interact with digital materials from the features of their eye movements, while qualitative approaches are able to explain why the interactions occur from the perspective of learners' perceptions. The combined method could help users rethink their prior learning behaviors. For example, the eye-tracking device provides the areas of interest, but why these areas are formed remains unknown. It may be attributed to various reasons such as learner interest, confusion, and so on, which requires further investigation to connect the eye movements to the specific reasons generated. Stark et al. (2018) applied the think-aloud approach to explore gaze patterns generated by eye-tracking, which supported the reliability of combining both methods to understand the deep cognitive processes. Although the eye tracking technique has generally been utilized in some small-scale experiments, with the development of deep learning in eye-tracking recognition, the PC camera could be used as an eye tracker, facilitating the popularization of the eye tracking technique in post-pandemic learning settings.

2.3. Stimulated recall in multimedia learning

Stimulated Recall has been extensively used to help learners to retrospect their learning behavior through the stimulus, such as recorded audios and videos captured in physical classrooms (Yuan & Lee, 2014). SR was developed based on the assumption that internal activities could be verbalized from the observed external real-world events. It has considerable potential to investigate studying cognitive strategies and learning processes (Geiger et al., 2016). Mackey and Gass (2016) also suggested that SR is an effective way to recognize learners' perceptions, their interpretation of events, and their thinking at a particular point. Although SR is widely used in physical contexts (Gazdag et al., 2019), it has been less explored and employed in remote learning settings, not to mention during the outbreak of a pandemic. The successful implementation of SR in online contexts may rely on the following two factors.

One factor is the stimulus captured from online learning behaviors, and the other is the retrospective strategy adapted to remote contexts. Although students' interactive behaviors could be recorded by video or audio, their interactions with multimedia are difficult to capture and analyze in online contexts. Thus, it is an urgent requirement to explore stimuli reflecting human-computer interaction, and further to motivate learners' retrospection in massive remote learning in the post-pandemic situation. Recent research has begun to explore the comprehensive understanding of biofeedback (e.g., eye tracking and EEG) in SR in the multimedia learning context (Zhai et al., 2018a). It is suggested that eye movements are reliable stimuli to SR and that retrospection is able to explain the biofeedback in return. Besides, stimulated recall affects the reliability of the retrospective strategy; for example, some designed questions have requested students to recall their cognitive behavior, by which learners' metacognition could be aroused (Abdel Latif, 2019). However, the outbreak of the COVID-19 pandemic may hinder the implementation of the retrospective strategy. Many instructors are struggling with increased workloads online and are experiencing elevated levels of anxiety and stress, and thus may neglect conducting retrospective instruction (Schmidt-Crawford et al., 2021). Likewise, during the suspension of classroom teaching, students perceive weak interactions between the digital content, which may vitiate their intentions to recall prior cognitive behaviors (Hamdan et al., 2021).

Synthesizing the above research background, this study aimed to address the following two research questions:

- Does the employment of multi-mode stimuli of feedforward and eye movements in stimulated recall improve learners' metacognition?
- How were the multi-mode stimuli compared and contrasted to inspire learners' metacognition in using digital dictionaries?

3. Methodology

3.1. Participants

The participants recruited in this study were native Mandarin speakers who were international graduate students in a university located in the southern United States. The participants were selected based on the following three criteria: (1) all the participants were familiar with the usage of digital dictionaries, so that they were well versed in their functions. (2) Participants should have adjusted to normal visual acuity to allow the eye-tracking software to properly calibrate. (3) Participants must be willing to perform immediate stimulated recall interviews. A total of 48 international students were finally recruited, and their demographics are shown in Table 1. To show our appreciation for their participation, gifts were sent to the participants after the experiment.

Table 1. Demographic profile of the participants

Categories		Frequency	Percentages
Age(years)	1=20-25	14	29.1
	2=26-31	16	33.3
	3=>31	18	37.5
Gender	1=Male	22	45.8
	2=Female	26	54.2
Degree Program	1=Master	29	60.4
	2=PhD	19	39.6
Major	1=Science	16	33.3
	2=Social science	19	39.6
	3=Art	13	27.1
English proficiency (TOEFL Scoring)	1=70-80	1	2
	2=80-90	39	81.3
	3=90-100	7	14.6
	4>100	1	2

3.2. Selection of digital dictionaries and vocabulary

The digital dictionary was an adaptive experimental platform for this study. Firstly, different from live broadcast platforms, digital dictionaries as auxiliary learning tools are generally used for online autonomous learning, and students' metacognitive activities are especially required in this situation (Connor et al., 2019). During the epidemic, learning activities were mostly carried out in a highly self-regulated learning context, with instruction and supervision by teachers lacking. The aim of employing a digital dictionary as a representative multimedia learning platform in this study was to inspire learners' metacognition in such self-regulated learning. Secondly, information and layout overload have been observed in many digital dictionaries (Frankenberg-Garcia, 2012) which distracts users from obtaining information efficiently and achieving their learning goals (Gouws & Tarp, 2016). When facing massive amounts of content and information provided by digital dictionaries, improving learners' judgment and awareness of valid information has become an urgent concern (Niitemaa & Pietilä, 2018). Thirdly, digital dictionaries have a broad user base for varied learning purposes, such as language learning and information searching (Levy & Steel, 2015; Lew & De Schryver, 2014). Given their distinctive learning motivations and varied using habits, many users expressed their desire for individualized services from digital dictionaries (Bastos & Machado, 2016). The above concerns led to our application of a digital dictionary in this study to verify the general applicability of the developed SR approach.

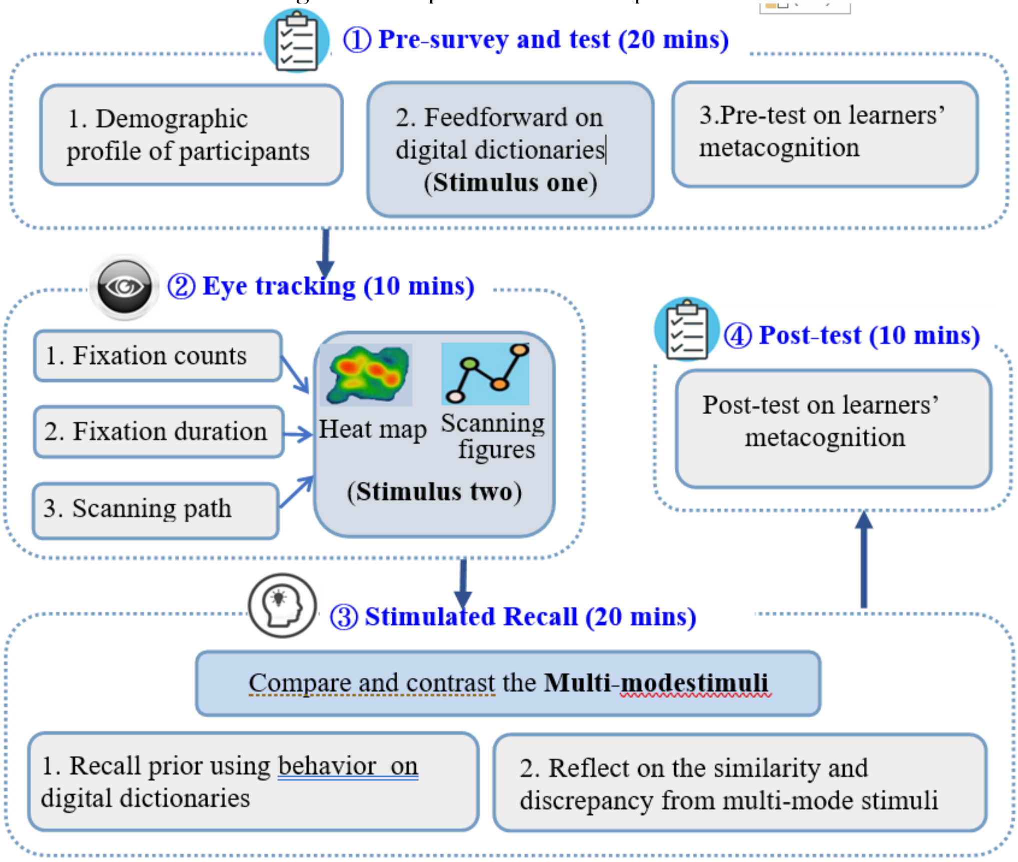
According to the criteria set by previous studies (Lew et al., 2013), the selected digital dictionaries should have similar functions and layout, including pronunciation, illustrations, definitions, phrases, synonyms/antonyms, and example sentences, so as to minimize the influence of functional distinction on users' perceptions during the experiment. Five digital dictionaries were selected according to the criteria mentioned for this study. Participants were surveyed to rank the dictionaries according to user experience. Finally, two of them were selected for this study: Youdao dictionary by Netease and Bing Dictionary by Microsoft, both of which were found to have a large user base.

Two criteria for selecting adaptive sample vocabulary in digital dictionary studies have been suggested: low frequency and polysemy (Dziemianko, 2015). By using low-frequency vocabulary, learners would focus on how to comprehensively utilize the functions of digital dictionaries to help them understand the vocabulary without being distracted by their prior knowledge of the vocabulary. According to the Corpus of Contemporary American English, low-frequency words are defined as those words that fall below the number 45,000 on a ranking of the most commonly used English words. Polysemous words were also suggested to be selected in digital dictionary studies, since polysemous words may maximize the functional display from the interface of digital dictionaries (Müller et al., 2015). According to the two criteria of vocabulary selection, two polysemous words, *tincture* and *sinew*, were selected by two professors specializing in linguistics. The researchers inputted the two words into the two digital dictionaries respectively, and took screenshots of the interfaces as source material for the eye tracking experiment.

3.3. Procedure and instruments

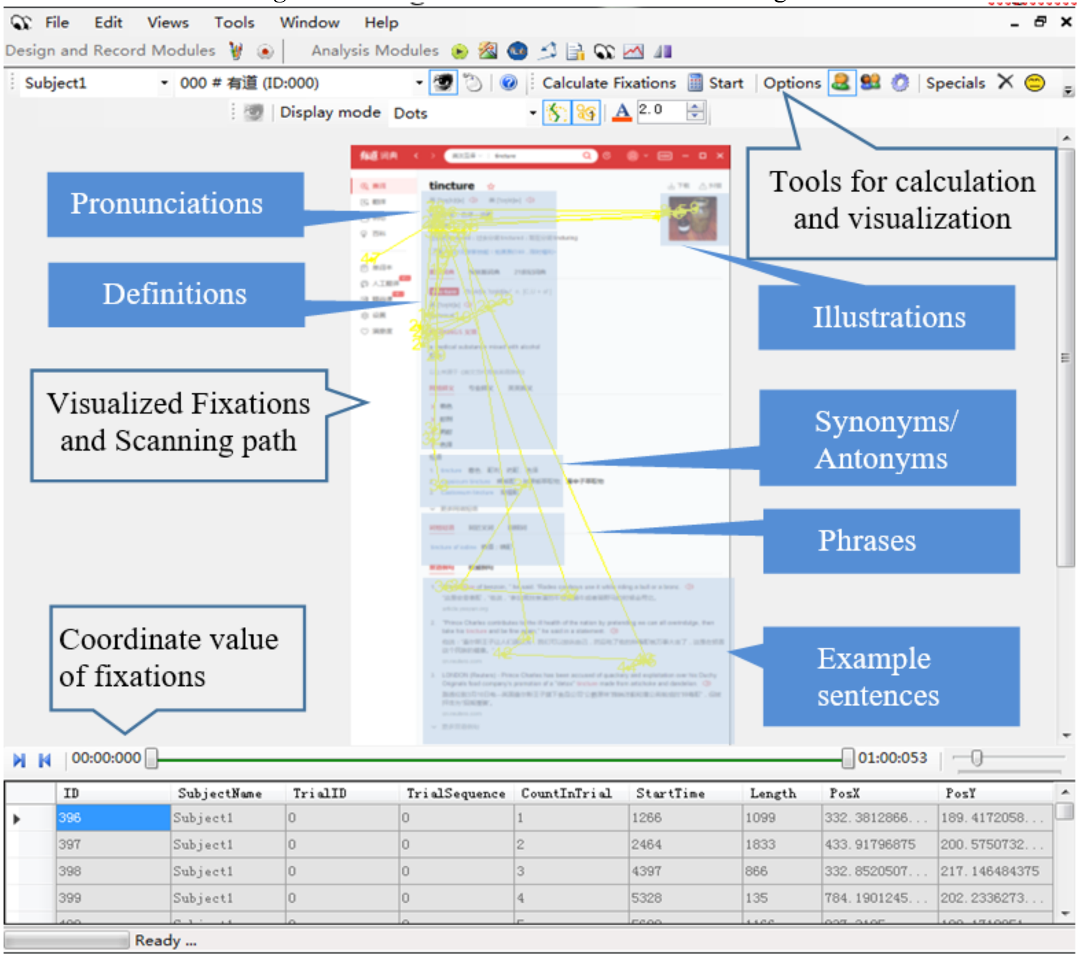
The procedure of the experiment shown in Figure 1 consists of four stages: a pre-survey and test, eye tracking, stimulated recall, and a post-test. Each participant spent around 60 minutes each time with help from an experienced teaching assistant. The experiment included a 20-minute pre-survey, 10-minute eye tracking, 20-minute stimulated recall, and a 10-minute post-test. In the first stage, a pre-survey was conducted to elicit participants' demographics, their feedforward on selected digital dictionaries, and a pre-test on metacognition. The pre-survey included demographics, feedforward, and the pre-test. According to the recommendations given in previous research on dictionaries (Collins, 2016; Frankenberg-Garcia, 2014), a 5-point scale (shown in Appendix 1), from *strongly disagree* to *strongly agree*, was employed to evaluate learners' feedforward on the perceived value of six typical functions in digital dictionaries (pronunciation, definitions, illustrations, phrases, synonyms/antonyms, and example sentences). Moreover, to examine students' metacognitive skills in the pre-post tests, a developed metacognitive scale, shown in Appendix 1, was adapted from Biasutti and Frate's research (2018).

Figure 1. The procedure of the experiments



The *Eye Tribe* eye-tracking device was employed in the second stage of the experiment. Having a reliable sampling rate from 30 Hz and 60 Hz mode, it is a reliable and adaptable tracker to capture learners' eye movements. Two open-access supporting software packages, Eyeproof and Ogama, were utilized to visualize the captured database displayed by heatmap and scanning paths (shown in Figure 2). The experiment was carried out in a laboratory with sound insulation. The participants were guided and acquainted with the equipment, procedures and the purpose of the eyetracking experiment, followed by signing the release form, granting permission to record their actions and comments. The experiment was conducted twice: once with the Youdao dictionary and once with the Bing dictionary. The Youdao dictionary was employed to display the interface of the selected vocabulary *tincture* the first time, and two days later, the Bing dictionary was used to present the interface of the other chosen vocabulary *sinew*. The experiments were conducted in one-by-one settings, since only one eye-tracker was utilized in this study. To minimize the interruption caused by the experiments, the researchers conducted the calibration of the eye movements by adjusting their head gesture, and helped participants get access to the test by a sample page, which could be completed in 5 minutes, so that the data collected in the first 5 minutes were discarded before participants got ready for the formal test.

Figure 2. The areas of interest and interface of Ogama



In the third stage, learners' feedforward and eye movements were presented to them as multi-mode stimuli in SR activities. On the basis of Cherrington and Loveridge's research (2014), the current research employed two-stepwise open questions, including recalling the original event and corrective feedback in this stage, by which learners not only reflected on how they used the dictionaries, but compared and contrasted the similarity and discrepancy between the feedforward and their displayed eye movements. Specifically, we asked (1) *What did you think according to your fixations and scanning paths;* (2) *Are there some conclusions by comparing eye movements and feedforward? What did you think of them?* To ensure the validity of recalling previous learning behaviors, previous studies have suggested that the SR interview should be conducted as soon as the experiment ends (Lyle, 2003). All the participants were interviewed by a teaching assistant approximately 20 minutes after the eye-tracking experiments were completed, and the SR interviews were recorded. After the SR, none of the participants reported any distraction due to the eye-tracker and SR interviews.

3.4. Data analysis

To test the effect of the SR with multi-mode stimuli on learners' metacognition, the current research employed the normal distribution and paired-samples *t* test by SPSS 19.0. Additionally, eye movements were evaluated by descriptive analysis and the Lag Sequential Analysis (LSA). The supporting software Ogama could generate the fixation counts and duration with adjustable criteria, and we set fixation at 200 ms in this study. Besides, to explore learners' scanning behaviors, the software GSEQ 5.1 was employed in this study to conduct the lag sequential analysis. According to the timed-event sequential data generated by Ogama, six events of the scanning paths, including *Pronunciation*, *Definitions*, *Illustrations*, *Phrases*, *Synonyms/Antonyms* and *Example Sentences*, were coded, which was followed by the implementation of the algorithms for computing inter-observer agreement.

4. Results

4.1. The pre-post tests of metacognition

In order to respond to the first research question, a *t* test was employed to measure learners' metacognitive skills according to the scoring of their pre-post tests. Firstly, the *t* value of the Kolmogorov-Smirnov (K-S) test shown in Table 2 scored 0.2 and 0.79 in the pre-test and post-test, which indicated that the data of metacognitive skills were normally distributed and qualified for the *t* test. Additionally, the reported mean value in the post-test was 4.14, while the result of the mean value in the pre-test was 3.34 ($p < .001$). The *t*-test results showed that the participants' metacognitive skills were significantly improved by the SR with multi-mode stimuli. The standard deviations are 0.26 and 0.24 in the pretest and posttest respectively, which indicated that the metacognition scoring was representative among the participants.

Table 2. The *t*-test results of learners' pre-post tests on metacognitive skill

	t-test results of metacognition					
	N	Mean	*S.D.*	K-S test	*t*-value	*p*
Pre test	48	3.34	0.26	0.20 (sig)	14.40	< .001
Post test	48	4.14	0.24	0.79 (sig)		

Table 3. The descriptive analysis of learners' feedforward and eye movements

	Feedforward				Eye-tracking on AOI	
	Perceived value (PV)		Expected Sum (ES)		Fixation counts (FC)	
	Mean	*S.D.*	Mean	*S.D.*	Mean	*S.D.*
1.Pronunciation	4.19	0.61	2.42	0.50	5.15	0.97
2.Definitions	4.69	0.55	2.92	0.61	4.90	0.90
3.Illustrations	4.60	0.57	2.73	0.71	5.50	1.07
4.Phrases	4.46	0.58	3.81	0.70	2.33	0.88
5.Synonyms/Antonyms	4.13	0.67	3.13	0.64	2.06	0.93
6.Example Sentences	4.52	0.58	4.25	0.76	15.45	5.38
(the first two)					14.33	5.36

Note. AOI refers to the area of interest.

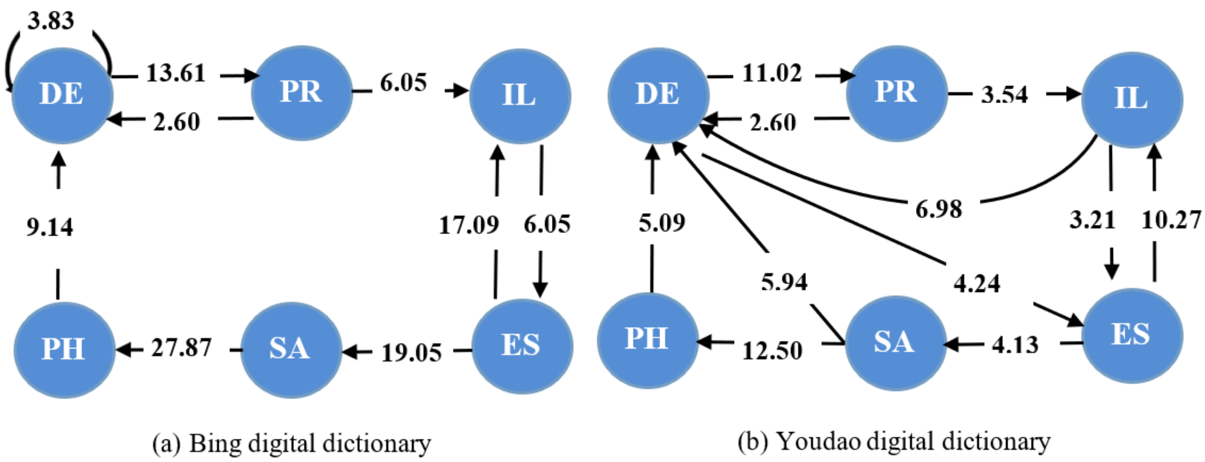
4.2. The comparison and contrast of feedforward and eye movements

In order to respond to the second research question, learners' eye movements and feedforward as multi-mode stimuli were compared and contrasted, by which learners retrospected why their eye movements were consistent or inconsistent with their prior knowledge, rather than merely recalling their prior behavior. Learners' average fixation counts (FC) and fixation duration (FD) of the two digital dictionaries were captured and generated from the supporting software Ogama. As shown in Table 3, the first three AOI on which fixation counts were mostly allocated were *Example Sentences* (FC = 15.45), *Illustrations* (FC = 5.5) and *Pronunciation* (FC = 5.15), followed by *Definitions* (FC = 4.9), *Phrases* (FC = 2.33) and *Synonyms/Antonyms* (FC = 2.06). According to the pre-survey of feedforward, the first three important functions in the digital dictionaries that the learners mostly emphasized were *Definition* (PV = 4.69), *Illustrations* (PV = 4.6), and *Example Sentences* (PV = 4.52), followed by *Phrases* (PV = 4.46), *Pronunciation* (PV = 4.19) and *Synonyms/Antonyms* (PV = 4.13). In terms of the expected sum, the first three expected functions in digital dictionaries were *Example Sentences* (ES = 4.25), *Phrases* (ES = 3.81) and *Synonyms/Antonyms* (ES = 3.13).

The above results show that learners' eye movements were partially in line with their feedforward. For example, according to the feedforward, the participants perceived a relatively higher value of using *Illustrations* and *Example Sentences* in the digital dictionaries, and their eye movements were found to be tallied with their feedforward. Likewise, *Phrases* and *Synonyms/Antonyms* were relatively less expected functionally, and the corresponding eye movements were less focused on these AOIs, which was consistent with their prior feedforward. However, as shown in Table 3 and Appendix 2, some discrepancies between learners' eye movements and their feedforward existed as well. For example, the sum of *Example Sentences* learners previously expected was 4.52 in their feedforward, while their concentrations were mainly allocated on the first two *Example Sentences* (14.33 out of 15.45 in fixation counts, and 12858.74 ms out of 13207.44 ms in fixation duration), which indicated that participants' cognitive load restricted their concentration on the rest of *Example Sentences*. Likewise, learners expected a relatively higher sum of phrases (*ES* = 3.81) and *Synonyms/ Antonyms* (*ES* = 3.13), whereas the fixation durations in these AOIs were both less than 900 ms, and the fixation counts were relatively less than that of other AOIs. (5) Besides, learners may have underestimated the value of *Pronunciation* (*PV* = 4.19, *ES* = 2.42), while relatively higher fixation counts and fixation durations were allocated in the AOI.

To further explain learners' eye movements when they used the selected digital dictionaries, Lag Sequential Analysis was employed. Six AOIs, *Pronunciation*, *Definitions*, *Illustrations*, *Phrases*, *Synonyms/Antonyms*, and *Example Sentences*, were coded in GSEQ 5.1. The eye movement data were generated from Ogama first, followed by the time sequence analysis. As shown in Figure 3, the arrows refer to the sequences learners performed from one event to another, and the coefficients predicted the correlations among the six events. The following observations in Figure 3 showed that: (1) The scan path started from *Definitions*, then moved to other events, and back to end on *Definitions*. The result indicated that the learning goal from the digital dictionaries was understanding the definitions of the selected vocabulary. (2) Their eyes moved between the *Definitions* and *Pronunciation* (z = 13.61 and 2.60 in Figure 3a. and z = 11.02 and 2.60 in Figure 3b), which suggested the assumption that users constantly tried to make connections between definitions and other multimedia information to help them understand the usage of the vocabulary. Besides, there was a set of iterative scanning behaviors between *Example Sentences* and *Illustrations* (z = 6.05 and 17.09 in Figure 3a. and z = 10.27 and 3.21 in Figure 3b), which shows that the learners tried to make illustrations connect to some specific social context.

Figure 3. Lag sequential analysis on learners' scanning behaviors



Note. DE, PR, IL, ES, SA, PH are the abbreviations of Definitions, Pronunciation, Illustrations, Example sentences, Synonyms/Antonyms, and Phrases respectively.

5. Findings and discussion

This study developed the stimulated recall approach with multi-mode stimuli to improve learners' metacognitive skills when they used digital dictionaries. Pre-post scale tests were conducted to answer the first research question, and the results showed a significant effect of multi-mode stimuli on participants' metacognitive skills. To further explain the driving mechanism, the similarities and discrepancies between feedforward and eye movements stimulated learners' metacognition as follows.

From the perspective of fixation, when learners recollected their eye movements, their cognitive behavior moved to the stage of extraneous processing and began to reorganize the material in the digital dictionaries.

Furthermore, when learners found that their eye movements were in line with their feedforward, they were encouraged to further confirm their assumptions with prior experiences, where their metacognition occurred in the stage of generative processing. For example, when students found that their fixation, consistent with feedforward, was focused on *Illustration*, they may have been involved in some real-world contexts. Two typical responses were generated from the SR interviews: (1) one is that students preferred the illustrations selected close to their previous experience and knowledge; (2) the other is that the illustrations should be located close to the definitions, which would help to spatially reduce the visual load. The findings are in line with prior research which found that learners were interested in connecting the information to their life experiences when using illustrations in digital language learning tools (Huang et al., 2012). Besides, although presented verbally, the Example Sentences provide a situational context for learners to picture activities from the sentences that help them understand the vocabulary (Huang et al., 2016). According to the SR interview, participants claimed the validity of picturing a specific contextual image in their minds, which enhanced their access to the application of the selected words. Students realized that they could activate their imagination of verbal material, which aroused their metacognitive skills of transforming media presentation. Likewise, synonyms and antonyms were relatively less noticed, which was inconsistent with their feedforward. Participants in the SR interview reported that although synonyms and antonyms of the selected words are useful to know, they were not involved in their initial learning goals. Participants realized that goal orientation is a keen factor in effective digital learning.

From the view of scanning path, definitions as an area of interest are focused. The lag sequential analysis on learners' scanning behaviors showed the significant sequential connection between definitions and some other multimedia elements, such as definitions and pronunciation. The dual coding assumption of CTML indicates that the dual channels, visual/pictorial and auditory/verbal, take effect simultaneously in the human-multimedia interaction system (Chen et al., 2017). There may be an interchange between the two channels in some situations, where users are capable of constructing their accordant psychological representations, which has been proven in the domain of vocabulary learning (Sadoski, 2005). Likewise, the significant sequential path between definitions and illustrations, shown in Figure 3(b), indicated that learners tried to recall their prior experience from the illustrations to gain access to the application of the selected words. The above retrospection may have aroused their metacognition on cross utilization of multiple elements in digital dictionaries to enhance their understanding of the unknown information. According to the SR interviews, participants concentrated on the definitions due to their confusion about which was the core and original meaning as well as the use frequency of the selected words, when polysemy was found.

Interestingly, some discrepancies between participants' feedforward and eye movements also existed, which helped learners activate their metacognition from conflict experiences (Questienne et al., 2018). Firstly, the data indicated that participants looked at the pronunciation section more often and for a longer period of time than they reported. According to the SR interviews, participants reflected on their desires for different voices, such as adult /child's voice, female/male's voice or young/older person's voice, to stimulate their auditory sense and help them remember the correct pronunciation of a word. This finding is consistent with the dual coding channels assumption. When physical representation and sensory representation are both shown, users' attention actively interacted between two modes in the multimedia learning context (Mayer, 2002). Participants' metacognition was generated by selecting their preferred pronunciation in order to meet their individualized learning requirements. Secondly, participants voiced their expectation of more example sentences. At the same time, focus was observed to be mainly on the first two example sentences, which may be due to the fact that some excessive information in the example sentences may have distracted their attention and increased their cognitive load. Some responses from the SR interview reflected that they might have overestimated their capacity, and their primary learning goal should be focused and split into specific sub-goals rather than remaining a desired but unachievable goal. Their psychological confusion may have generally occurred due to the conflicts between their feedforward and their eye movements, which awakened their metacognition that the resourceful multimedia in digital dictionaries may exceed their cognitive system's processing capacity.

6. Implications

6.1. Theoretical implications

CTML emphasizes the importance of metacognition for digital learning, and proposes finding a cuing factor as a stimulus to inspire learners' metacognitive skills (Mudrick et al., 2019), while less research has theoretically constructed a principle for it. The theoretical implication of this study contributes to exploring a principle to improve metacognition with multi-mode stimuli in digital learning settings. We recommend that multi-mode stimuli be designed and developed in SR to investigate learners' metacognition. To help learners recall their prior

behaviors, stimulated recall in previous studies employed either surveys or recorded materials such as photos, audios, and videos as stimuli. In this situation, learners most likely tend to recall the knowledge presented before, but may ignore retrospecting their previous cognitive behaviors and learning strategies.
Multi-mode stimuli are helpful for arousing learners' metacognition from two channels. For the first channel, the multi-mode stimuli mutually explain the driving mechanism of the behaviors. A single stimulus is weak in explaining the reason why behaviors happen. For example, eye tracking can tell where eyes linger but cannot explain why. The multi-mode stimuli could improve the deficiency of a single stimulus, which can enhance learners' deep understanding of the driving mechanism of learning behavior and strategies. The other channel is generating metacognition from conflict. Conflict occurs because individuals' cognitive capacity may consistently vary throughout their lives, and their expectations of cognitive capacity may be relatively hysteretic to their real behaviors. When discrepancy occurred among multi-mode stimuli, learners would instinctively retrospect the reasons caused in terms of the adaptation between prior cognitive mode and real learning behaviors.

The effect of COVID-19 has gradually impelled blend learning as an important learning channel. In the context of blend learning, the data sources are not only multivariate, but also multistage. Learners are encouraged to make comparison and contrast of these stimuli, by which their metacognition are expected to be inspired. First of all, with the assistance of information technology, the multi-mode learning behaviors and perceptions are accessible in the blend learning context. For example, the PC camera with the assistance of deep learning algorithm could be utilized as an eye tracker to capture learners' eye interaction with online material (SM et al., 2021), while their log data are also accessible from intelligent tutor system. Additionally, the longitudinal data instead of cross-section data are suggested to be employed as stimuli in this approach. Learners could much more effectively focus and reflect on a specific learning procedure between the stage of longitudinal stimuli collected, rather than recall their behaviors in general.

6.2. Practical implications

The current research has a series of practical implications for the modification and redesigning of digital dictionaries. Firstly, it is suggested that diversified pronunciation recordings, such as using a child's voice, be provided to meet learners' individual preferences. Current digital dictionaries typically only offer one or two pronunciation voices. It may be beneficial to provide a broad selection of pronunciation voices, such as the voices of male, female, older adult and young speakers to meet learners' personalized preferential treatment demands.

Second, the number of example sentences selected in digital dictionaries should be taken into consideration. According to eye-tracking data, it is suggested that cognitive overload is a central challenge in the design of digital dictionaries. Therefore, users should be able to toggle the maximum number of example sentences that appear for each entry according to their cognitive load. In addition, example sentences should be diversified to cover equivalent contexts and definitions. Even when a significant number of example sentences are presented, many of them are only for high-frequency definitions, ignoring the polysemy. The possible reason may be that example sentences were selected automatically according to search engines, which were the most frequently used but not always the most suitable. Therefore, the definition, illustration, and example should be consistent with each other.

Third, it is suggested that illustrations be spatially close to the corresponding definitions and example sentences. Based on the contiguity principle in CTML, learners' cognitive load is reduced when text and graphics are tightly spatially integrated, rather than presented separately. The illustrations selected should (1) have a relation with daily life, and (2) relate to the definitions and example sentences. Participants claimed that pictures related to their routine contributed more to their learning efficiency and drastically lowered their cognitive load. According to CTML, generative processing happens when learners actively integrate prior knowledge into working memory. Selective illustrations could facilitate learners' building of connections between words, definitions, example sentences, and pictures. They could successfully extract previous knowledge from long-term memory, then integrate the processed information with prior knowledge.

Finally, the digital dictionaries should help users to engage in the construction and modification of the dictionary content or user interface. Users showed strong interest in participating in providing feedback such as the selection criteria of the illustrations, and modifications of definition. Digital dictionaries, therefore, could be designed as an open-access or semi-open system for users' deep involvement. Digital dictionaries could not only be used as a tool for information searching, but could also help to construct a collaborative and creative learning context.

7. Limitations

Although a rigorous validation procedure was implemented to investigate learners' general using behavior and cognitive processes while using digital dictionaries, this research still suffers from some limitations. Firstly, only 48 participants from one university were recruited for this study, all of whom were native Mandarin speakers. To deepen our understanding of the individualized requirements of digital dictionaries, more consideration should be given to students' varied personalities, cultural backgrounds, language levels, technology self-efficacy, and so on. Secondly, the measurement of learners' metacognition was evaluated in general, and the results of each dimension of metacognition were not involved in this research. Thirdly, only digital dictionaries were selected as the research platform; there are, however, many other multimedia learning tools which could be involved in future studies. Finally, due to the limited function of the device and software, only the images of the final entries of digital dictionaries were included and studied, and so there will be further exploration of the synchronous influence of both visual and audio stimuli.

References

Abdel Latif, M. M. (2019). Using think-aloud protocols and interviews in investigating writers' composing processes: Combining concurrent and retrospective data. *International Journal of Research & Method in Education*, *42*(2), 111-123.

Abdullah, M. N. L. Y. (2020). The Influence of self-regulation processes on metacognition in a virtual learning environment. *Educational Studies*, *46*(1), 1-17.

Antonietti, A., Colombo, B., & Di Nuzzo, C. (2015). Metacognition in self-regulated multimedia learning: Integrating behavioural, psychophysiological and introspective measures. *Learning, Media and Technology*, *40*(2), 187-209.

Bastos, H. P., & Machado, G. P. (2016). Dictionaries on Smartphones: Learners' assessment of features and potential of dictionary apps as pedagogical tools. In *New Advances in Information Systems and Technologies* (pp. 143-156). Springer, Cham.

Biasutti, M., & Frate, S. (2018). Group metacognition in online collaborative learning: Validity and reliability of the group metacognition scale (GMS). *Educational Technology Research and Development, 66*(6), 1321-1338.

Chen, Z. H., Chen, S. Y., & Chien, C. H. (2017). Students' reactions to different levels of game scenarios: A Cognitive style approach. *Educational Technology & Society, 20*(4), 69-77.

Chakma, U., Li, B., & Kabuhung, G. (2021). Creating online metacognitive spaces: Graduate research writing during the COVID-19 pandemic. *Issues in Educational Research, 31*(1), 37-55.

Cherrington, S., & Loveridge, J. (2014). Using video to promote early childhood teachers' thinking and reflection. *Teaching and Teacher Education, 41*, 42-51.

Collins, J. B. (2016). Changes in electronic dictionary usage patterns in the age of free online dictionaries: Implications for vocabulary acquisition. *APU Journal of Language Research, 1*, 36-49.

Connor, C. M., Day, S. L., Zargar, E., Wood, T. S., Taylor, K. S., Jones, M. R., & Hwang, J. K. (2019). Building word knowledge, learning strategies, and metacognition with the Word-Knowledge e-Book. *Computers & Education, 128*, 284-311.

Cortina, K. S., Miller, K. F., McKenzie, R., & Epstein, A. (2015). Where low and high inference data converge: Validation of CLASS assessment of mathematics instruction using mobile eye tracking with expert and novice teachers. *International Journal of Science and Mathematics Education, 13*(2), 389-403.

Dulamă, M. E., & Ilovan, O. R. (2016). How powerful is feedforward in university education? A Case study in Romanian geography education on increasing learning efficiency. *Educational Sciences: Theory & Practice, 16*(3), 827-848.

Dunlosky, J., & Metcalfe, J. (2008). *Metacognition*. Sage Publications.

Dunlosky, J. (2005). Why does rereading improve metacomprehension accuracy? Evaluating the levels-of-disruption hypothesis for the rereading effect. *Discourse Processes, 40*(1), 37-55.

Dziemianko, A. (2015). Colours in online dictionaries: A Case of functional labels. *International Journal of Lexicography, 28*(1), 27-61.

Edwards, O. V. (2010). *The Effect of goal orientation of attention, learning, and metacognitive awareness* (Unpublished doctorial dissertations). University of Nevada, Las Vegas.

Frankenberg-Garcia, A. (2012). Learners' use of corpus examples. *International Journal of Lexicography, 25*(3), 273-296.

Frankenberg-Garcia, A. (2014). The Use of corpus examples for language comprehension and production. *ReCALL, 26*(2), 128-146.

Frith, C. D. (2012). The Role of metacognition in human social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1599), 2213-2223.

Gazdag, E., Nagy, K., & Sziväk, J. (2019). “I Spy with My Little Eyes…” The use of video stimulated recall methodology in teacher training–The exploration of aims, goals and methodological characteristics of VSR methodology through systematic literature review. *International Journal of Educational Research, 95*, 60-75.

Geiger, V., Muir, T., & Lamb, J. (2016). Video-stimulated recall as a catalyst for teacher professional learning. *Journal of Mathematics Teacher Education, 19*(5), 457-475.

Gouws, R. H., & Tarp, S. (2016). Information overload and data overload in lexicography. *International Journal of Lexicography, 30*(4), 389-415.

Hamdan, K. M., Al-Bashaireh, A. M., Zahran, Z., Al-Daghestani, A., AL-Habashneh, S., & Shaheen, A. M. (2021). University students’ interaction, Internet self-efficacy, self-regulation and satisfaction with online education during pandemic crises of COVID-19 (SARS-CoV-2). *International Journal of Educational Management, 35*(3), 713-725.

Horská, E., Nagyová, Ľ., Šedík, P., Kiełbasa, B., & Krasnodębski, A. (2020). Evaluation of cheese packaging graphics design using eye tracking and consumer survey research. *Journal of Management Cases*, *22*(2), 38-45.

Huang, C. S., Yang, S. J., Chiang, T. H., & Su, A. (2016). Effects of situated mobile learning approach on learning motivation and performance of EFL students. *Educational Technology & Society, 19*(1), 263-276.

Huang, Y.-M., Huang, Y.-M., Huang, S.-H., & Lin, Y.-T. (2012). A Ebiquitous English vocabulary learning system: Evidence of active/passive attitudes vs. usefulness/ease-of-use. *Computers & Education, 58*(1), 273-282.

Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., Lee, M.-H., Chiou, G.-L., Liang, J.-C., & Tsai, C.-C. (2013). A Review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review, 10*, 90-115.

Leszkowicz, M. (2011). “Keeping up with the Joneses.” A Sociological content analysis of advertising catalogues with the eye-tracking method. *Qualitative Sociology Review, 7*(2), 85-100.

Levy, M., & Steel, C. (2015). Language learner perspectives on the functionality and use of electronic language dictionaries. *ReCALL, 27*(2), 177-196.

Lew, R., Grzelak, M., & Leszkowicz, M. (2013). How dictionary users choose senses in bilingual dictionary entries: An Eye-tracking study. *Lexikos, 23*(1), 228-254.

Lew, R., & De Schryver, G. M. (2014). Dictionary users in the digital revolution. *International Journal of Lexicography*, *27*(4), 341-359.

Lorigo, L., Haridasan, M., Brynjarsdóttir, H., Xia, L., Joachims, T., Gay, G., Granka, L., Pellacini, F., & Pan, B. (2008). Eye tracking and online search: Lessons learned and challenges ahead. *Journal of the American Society for Information Science and Technology, 59*(7), 1041-1052.

Luo, L., Kiewra, K. A., Peteranetz, M. S., & Flanigan, A. E. (2017). Using eye-tracking technology to understand how graphic organizers aid student learning. In *Eye-tracking Technology Applications in Educational Research* (pp. 220-238). IGI Global. https://doi.org/10.4018/978-1-5225-1005-5.ch011

Lyle, J. (2003). Stimulated recall: A Report on its use in naturalistic research. *British Educational Research Journal, 29*(6), 861-878.

Mackey, A., & Gass, S. M. (2016). *Stimulated recall methodology in applied linguistics and L2 research*. Routledge.

Mayer, R. E. (2002). Multimedia learning. In *Psychology of learning and motivation* (Vol. 41, pp. 85-139). Academic Press.

Mayer, R. E. (2014). Multimedia instruction. In *Handbook of Research on Educational Communications and Technology* (pp. 385-399). Springer.

Meuwese, J. D., van Loon, A. M., Lamme, V. A., & Fahrenfort, J. J. (2014). The Subjective experience of object recognition: Comparing metacognition for object detection and object categorization. *Attention, Perception, & Psychophysics, 76*(4), 1057-1068.

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review, 19*(3), 309-326.

Mudrick, N. V., Azevedo, R., & Taub, M. (2019). Integrating metacognitive judgments and eye movements using sequential pattern mining to understand processes underlying multimedia learning. *Computers in Human Behavior, 96*, 223-234.

Nasirzade, S., & Nargesian, J. (2019). The Effect of metacognitive beliefs on academic well-being mediated by perfectionist students. *Journal of School Psychology, 8*(3), 177-195.

Niitemaa, M.-L., & Pietilä, P. (2018). Vocabulary skills and online dictionaries: A Study on EFL learners' receptive vocabulary knowledge and success in searching electronic sources for information. *Journal of Language Teaching and Research, 9*(3), 453-462.

Park, B., Knörzer, L., Plass, J. L., & Brünken, R. (2015). Emotional design and positive emotions in multimedia learning: An Eyetracking study on the use of anthropomorphisms. *Computers & Education, 86*, 30-42.

Parlan, P., & Rahayu, S. (2021, March). Students' higher order thinking skills (HOTS) in metacognitive learning strategy. In *AIP Conference Proceedings* (Vol. 2330, No. 1, pp. 020035). AIP Publishing LLC. https://doi.org/10.1063/5.0043150

Questienne, L., Van Opstal, F., van Dijck, J. P., & Gevers, W. (2018). Metacognition and cognitive control: Behavioural adaptation requires conflict experience. *Quarterly Journal of Experimental Psychology, 71*(2), 411-423.

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457-1506.

Renkl, A., Berthold, K., Grosse, C. S., & Schwonke, R. (2013). Making better use of multiple representations: How fostering metacognition can help. In *International Handbook of Metacognition and Learning Technologies* (pp. 397-408). Springer. https://doi.org/10.1007/978-1-4419-5546-3_26

Rivers, M. L. (2020). Metacognition about practice testing: A Review of learners' beliefs, monitoring, and control of test-enhanced learning. *Educational Psychology Review, 33*(3), 823-862. https://doi.org/10.1007/s10648-020-09578-2

Rollwage, M., Dolan, R. J., & Fleming, S. M. (2018). Metacognitive failure as a feature of those holding radical beliefs. *Current Biology, 28*(24), 4014-4021.

Sadoski, M. (2005). A Dual coding view of vocabulary learning. *Reading & Writing Quarterly*, *21*(3), 221-238.

Schmidt-Crawford, D. A., Thompson, A. D., & Lindstrom, D. L. (2021). Condolences and congratulations: COVID-19 pressures on higher education faculty. *Journal of Digital Learning in Teacher Education*, *37*(2), 84-85.

SM, H. K., Pradyumna, G., Aishwarya, B., & Gayathri, C. (2021). Development of personal identification number authorization algorithm using real-time eye tracking & dynamic keypad generation. In *2021 6th International Conference for Convergence in Technology (I2CT)* (pp. 1-6). IEEE.

Stark, L., Brünken, R., & Park, B. (2018). Emotional text design in multimedia learning: A Mixed-methods study using eye tracking. *Computers & Education, 120*, 185-196.

Taub, M., & Azevedo, R. (2019). How does prior knowledge influence eye fixations and sequences of cognitive and metacognitive SRL processes during learning with an intelligent tutoring system*? International Journal of Artificial Intelligence in Education, 29*(1), 1-28.

Wang, C.-Y., Tsai, M.-J., & Tsai, C.-C. (2016). Multimedia recipe reading: Predicting learning outcomes and diagnosing cooking interest using eye-tracking measures. *Computers in Human Behavior, 62*, 9-18.

Wang, J., Antonenko, P., Celepkolu, M., Jimenez, Y., Fieldman, E., & Fieldman, A. (2019). Exploring relationships between eye tracking and traditional usability testing data. *International Journal of Human–Computer Interaction, 35*(6), 483-494.

Yuan, R., & Lee, I. (2014). Pre-service teachers' changing beliefs in the teaching practicum: Three cases in an EFL context. *System, 44,* 1-12.

Zhai, X., Fang, Q., Dong, Y., Wei, Z., Yuan, J., Cacciolatti, L., & Yang, Y. (2018a). The Effects of biofeedback-based stimulated recall on self-regulated online learning: A Gender and cognitive taxonomy perspective. *Journal of Computer Assisted Learning, 34*(6), 775-786.

Zhai, X., Meng, N., Yuan, J., Yang, Y., & Lin, L. (2018b). Exploring learners' cognitive behavior using e-dictionaries: An Eye-tracking approach. In *International Conference on Innovative Technologies and Learning* (pp. 165-171). Springer.

Zhang, R., & Zou, D. (2021). A State-of-the-art review of the modes and effectiveness of multimedia input for second and foreign language learning. *Computer Assisted Language Learning*, 1-27.

Appendix 1

The metacognition scale and the pre-survey on the perceptions of digital dictionaries

Items on metacognition	1	2	3	4	5
(1) When using digital dictionaries, I know my strengths as a learner.					
(2) When using digital dictionaries, I know how to select relevant information.					
(3) I know how to use the material in digital dictionaries.					
(4) I know how to organize new information in digital dictionaries.					
(5) When using digital dictionaries, I know how to connect new information with prior knowledge.					
(6) I can plan the activities when I use digital dictionaries.					
(7) When using digital dictionaries, I determine what the task requires.					
(8) When using digital dictionaries, I can select the appropriate functions.					
(9) When using digital dictionaries, I can identify the strategies depending on the task.					
(10) When using digital dictionaries, I organize my time depending on the task.					
(11) When using digital dictionaries, I modify my work according to other participants' suggestions.					
(12) I am used to asking questions to check my understanding when using digital dictionaries.					
(13) I check my approach to improve our outcomes when using digital dictionaries.					
(14) I improve my work with group processes when using digital dictionaries.					
(15) I detect and correct my errors when using digital dictionaries.					
(16) I make judgments on the difficulty of the task when using digital dictionaries.					
(17) I make judgments on the workload when using digital dictionaries.					
(18) I make judgments on the instruments when using digital dictionaries.					
(19) I make judgments on my learning outcomes when using digital dictionaries.					
(20) I make judgments on the teamwork process when using digital dictionaries.					
Items on feedforward in digital dictionaries	**1**	**2**	**3**	**4**	**5**
(1) I expect to use digital pronunciation to improve my learning when using digital dictionaries.					
(2) The explanatory definitions in the digital dictionaries are important to me.					
(3) The illustrations embodied in the digital dictionaries are helpful for learning.					
(4) I think the phrases are useful for learning when using digital dictionaries.					
(5) The synonyms/antonyms are important elements designed in digital dictionaries.					
(6) Example sentences benefit me a lot when I learn with digital dictionaries.					
	none	**1**	**2-3**	**4-5**	**>5**
(7) How many forms of the digital pronunciations do you suggest that digital dictionaries should offer?					
(8) How many explanatory definitions do you expect from digital dictionaries?					
(9) How many illustrations do you suggest should be presented in digital dictionaries?					
(10) How many phrases do you suggest should be provided by digital dictionaries?					
(11) How many synonyms/antonyms do you expect from digital dictionaries?					
(12) How many example sentences do you suggest that digital dictionaries should provide?					

Appendix 2

The descriptive analysis of learners' fixation duration on each AOI

	Fixation duration (ms) (FD)	
	Mean	*S.D.*
1.Pronunciation	4058.88	1206.08
2.Definitions	3687.67	1307.62
3.Illustrations	3721.90	1358.49
4.Phrases	865.79	433.90
5.Synonyms/Antonyms	688.48	382.33
6.Example Sentences	13207.44	5848.66
(the first two)	12858.74	5777.70

Note. The fixation duration was calculated in milliseconds. AOI refers to the area of interest.