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

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

Apr. 2020 Journal of Educational Technology & Society

vol.23 no.2

Journal of Educational Technology & Society

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

Volume 23 Issue 2
April 2020

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

Educational Technology & Society
An International Journal

Aims and Scope

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

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

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

Publisher
International Forum of Educational Technology & Society

Host
National Yunlin University of Science and Technology, Taiwan

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

Supporting Organizations
University of North Texas, USA
University of Piraeus, Greece

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

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

Guidelines for authors
Submissions are invited in the following categories:

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

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

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

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

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

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

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

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

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

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

Educational Technology & Society

Volume 23 Number 2 2020

Table of contents

Full Length Articles

Yoo, M., & Jin, S. (2020). Development and Evaluation of Learning Analytics Dashboards to Support Online Discussion Activities. *Educational Technology & Society*, *23* (2), 1–18.

Development and Evaluation of Learning Analytics Dashboards to Support Online Discussion Activities

Mina Yoo[1] and Sung-Hee Jin[2*]
[1]Kyungil University, Republic of Korea // [2]Hanbat National University, Republic of Korea // ymn@kiu.kr // shjin@hanbat.ac.kr
[*]Corresponding author

(Submitted January 20, 2020; Revised April 25, 2020; Accepted May 18, 2020)

ABSTRACT: Online discussion plays an increasingly significant role in asynchronous online learning environments. While previous attempts have been made to develop learning analytics dashboards to facilitate such discussions, most of these dashboards have been designed without reference to data or visualization techniques that have been proven to make online discussions more effective. This study identified the difficulties and inconveniences experienced by learners in online discussion activities and generated a set of visual design guidelines for overcoming them. Applying these guidelines, a set of learning analytics dashboards were developed and evaluated. The study was conducted according to prototyping methodology, which yielded five prototype dashboards that display information on participation, interaction, discussion content keywords, discussion message types, and the distribution of debate opinions, respectively. The developed dashboards were then revised and refined in a three-step process: (1) expert validation to verify that the dashboards complied with the visual guidelines and provided learners with the information they needed; (2) tests to identify usability problems, collect qualitative and quantitative data, and determine participant satisfaction; and (3) user experience evaluations to determine how learners and instructors perceived their interactions with the dashboards. Practical and empirical discussions are provided based on the results, which offer a valuable base of user experience data that can be used in future studies.

Keywords: Visual dashboard, Prototype development, Online discussion, Learning analytics, Prototyping methodology

1. Introduction

Online discussions are learning activities commonly used in online learning environments, such as MOOCs or flipped learning. Asynchronous online discussions are useful for analyzing and reflecting on discussion content, because they allow learners to repeatedly check for discussion opinions (Hara, Bonk, & Angeli, 2000) with no constraints of time or space, unlike traditional face-to-face class discussions (Harasim, 1993). Such conversations take place on online forum bulletin boards, where text-based discussion accumulates over time. As such, learners are required to exert a large amount of effort to understand the discussion's overall flow and respond appropriately (Wise, Zhao, & Hausknecht, 2013). Thus, discussion forums are difficult to understand and manage when a large number of students are discussing ideas (Vieira, Parsons, & Byrd, 2018).

Researchers have previously adopted a learning analytics perspective to analyze and visualize discussion activities in a dashboard format so as to support learners' understanding and monitoring of online discussion activities. There is evidence that dashboards may promote learning by providing learners with opportunities to monitor and reflect on their learning process (Verbert, Duval, Klerkx, Govaerts, & Santos, 2013). However, most dashboards are designed through ad-hoc processes rather than in consultation with the results of rigorous research. Moreover, the lack of research on appropriate visualization techniques for each data type leads to the creation of ineffective dashboards (Verbert et al., 2019), and many learning dashboards are implemented without conducting usability tests on learners (Bodily & Verbert, 2017). A successful implementation of learning analytics dashboards requires considering learners' needs (Ifenthaler, 2017).

Against this background, this study sought to develop learning analytics dashboards applicable to online discussions by exploring the educationally meaningful information in online discussion activities and applying research-based guidelines to visualize it in the most effective way. Our chosen learning analytics dashboards were revised and validated four ways using expert reviews, expert validations, usability tests, and user experience evaluations. The results of this study offer practical and empirical paths forward for the development of visual dashboards for online discussion activities.

2. Literature review

2.1. Online discussion and visual feedback information

Rapport (1991) defined online discussions as interactions in which learners exchange text-based messages in a virtual space in a many-to-many format. Discussion forums have been widely applied and have many benefits as a teaching and learning tool for both blended and online courses in numerous disciplines (An, Shin, & Lim, 2009). Compared to face-to-face contexts, online discussion environments can encourage wider learner participation because there are no time or space constraints (Buckley, 2011; Harasim, 2000). Furthermore, learners can improve their critical thinking skills and generate new knowledge through collaborations with other learners (Weinberger & Fischer, 2006; Hew & Cheung, 2011). Online discussion forums also allow learners to gather and organize relevant data on a particular topic before contributing to the discussion (Hara et al., 2000).

Despite these educational benefits, online discussion activities have certain educational limitations that must be overcome. For example, learners may have difficulty with behavioral regulation during participation activities, or with understanding the meaning or detailed aspects of written messages (Kehrwald, 2008). To avoid such pitfalls, previous research has adopted a learning analytics approach, seeking to help learners better understand discussion activities by visually representing a range of results.

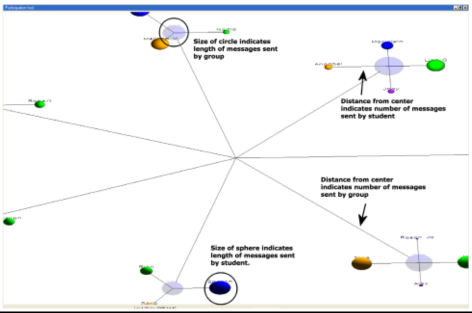
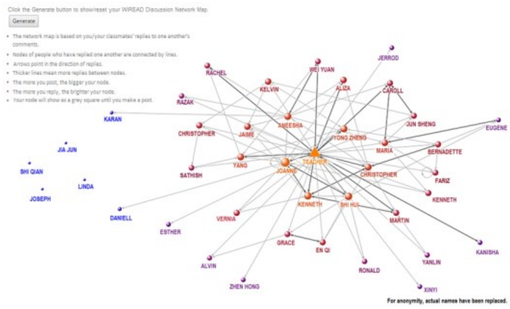
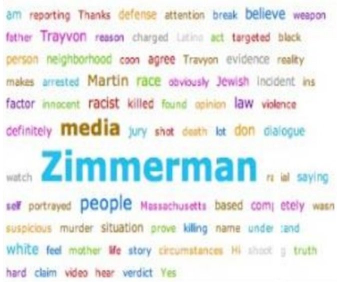
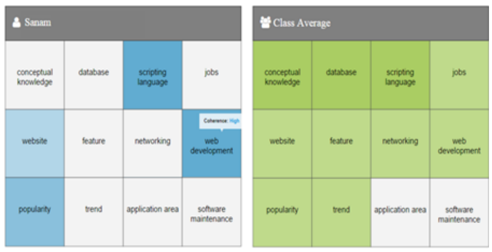
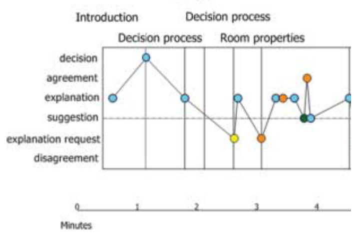
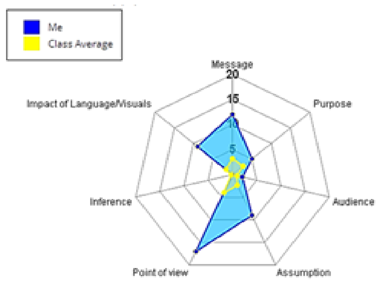
Following this approach, learners receive feedback on discussion activities through visualizations of the results of their participation, interaction, and discussion content analysis in individual or team discussions. Such visual feedback on the learning process and outcomes enables learners to objectively monitor their learning activities and understand the current state of the interactions among peer learners (Ferguson, 2012). When learning data is related to learning objectives and is able to track learners' progress, meaningful visual feedback can be created to enhance desired learning behaviors according to a process model (Verbert et al., 2013). As a result, learners may gain a better overview of discussion activities (*awareness*), reflect on their own activities (*self-reflection*), find their deficiencies (*sensemaking*), and change their learning behavior to compensate for these deficiencies (*impact*). To provide visual feedback on online discussion activities, it is important to identify the information that learners and instructors require (Yau, 2013). Due to the nature of online learning environments, however, some information is more difficult for learners and instructors to obtain than in face-to-face discussion activities. Table 1 shows the characteristics of online discussion activities proposed in previous research, learners' difficulties in online discussion forums, and the items that should be presented as visual feedback information.

Table 1. Visual feedback information focusing on the difficulties in online discussion

Characteristics of online discussion environments	Difficulties in online discussion	Visual feedback information
Independent	• Difficulty in making an objective self-evaluation of how to engage actively in discussion activities	(1) Online discussion participation (Hatala, Beheshitha, & Gasevic, 2016; Tan, Koh, Jonathan, & Yang, 2017)
Text-based conversations	• Difficulty in understanding interactions among learners • Difficulty in understanding the overall discussion	(2) Interaction among learners (Dawson, Bakharia, & Heathcote, 2010) (3) Keywords (Ali, Hatala, Gašević, & Jovanović, 2012) (4) Message types (Pallotta & Delmonte, 2011)

Since each student accesses the platform asynchronously and participates in online discussion activities independently, it is difficult for learners to observe other learners' discussion activities and so make objective self-assessments regarding their relative level of active participation. Yet, this participation as a basic learning behavior is the most important predictor of the educational effectiveness of online discussion (Jin, Yoo, & Kim, 2015). As seen in Table 2, previous studies have visualized individual learners' participation levels in online discussions either over time or in comparison with other learners. For example, Govaerts, Verbert, Duval, and Pardo (2012) assessed learners' participation over time, represented as a line graph, allowing each individual to compare their participation with other learners. By contrast, Jin et al. (2015) represented the participation levels of individual learners and teams using ten color codes, making it easier for them to grasp their participation information at a glance.

Table 2. Learning analytics dashboards on visual feedback information

Visual information	Online discussion dashboards	
Participation	Individual learner participation by date (Govaerts et al., 2012) 	Individual and team participation using ten color codes (Jin et al., 2015) 
Interaction	Interaction between teams (Janssen, Erkens, Kanselaar, & Jaspers, 2007) 	Interaction between individuals (Tan et al., 2017) 
Keywords	Frequency of words used by learners (Murray, 2014) 	Relevance between keywords suggested by instructors and words written by learners (Hatala et al., 2016) 
Message types	Learners' opinion analysis according to argumentative categories (Pallotta & Delmonte, 2011) 	Type of message selected by learners (Tan et al., 2017) 

Since online discussion activities are usually text-based conversations that accumulate messages in a bulletin board format, it is often difficult to grasp the relationships of interaction among learners. In this regard, some studies have provided visualizations of these interactions among individual learners and teams (Janssen, Erkens, Kanselaar, & Jaspers, 2007; Tan et al., 2017), representing activities related to writing their own opinions, reading the opinions of others, and posting comments or replies. Visual feedback on learner interactions can improve participants' presence and co-presence in online discussions by helping students clearly recognize their own and other learners' discussion activities (Lambropoulos, Faulkner, & Culwin, 2012). Sociograms composed of nodes and links are widely used to visualize these interactions. Expressing the learner as a node and the relationship among learners as a link, they can easily provide information on who is leading the discussion and who is being excluded.

Additional studies have extracted and presented keywords or analyzed the types of messages written by learners to help them better understand discussion topics and opinions. In these, key or common words used by learners are represented in word clouds (Murray, 2014) or by visualizing the semantic relationships between words (Sack, 2000). For example, Hatala et al. (2016) determined keywords related to the discussion topic and presented the degree to which those keywords were used to suit the content of the learners' discussion opinions, using four color-coded levels. In terms of analysis, Pallotta and Delmonte (2011) evaluated and visualized the types of messages written by learners over time according to argumentative categories, while Tan et al. (2017) offered learners seven critical lenses, required that they select one to write about, and portrayed the results in a radial graph.

2.2. Visual design guidelines for learning analytics dashboards

Visual design guidelines are essential for the effective design and development of learning analytics dashboards for online discussion forums. They can serve as a tool for determining how and what to visualize in online discussion activities (Yau, 2013). Learning analytics dashboards provide a visual representation of the learning process, giving both learners and instructors an interactive aggregation of individual and group goals, tasks, connections and achievements in real time (Alabi, Code, & Irvine, 2013). These dashboards enhance learning by providing feedback, especially behavioral process-oriented feedback to support students' learning regulation (Sedrakyan, Malmberg, Verbert, Järvelä, & Kirschner, 2020). The visual design guideline presented in this study considers the relationship between visual dashboard design and learning analytics to provide process-oriented feedback that supports behavioral regulation in online discussion activities. Previous studies proposed visual guidelines for presenting the results of learning analytics that can be summarized as traceability, comparability, implicity, and overview plus detail.

Traceability refers to visualizing and portraying analysis results in the order that learning activities continuously occur, covering past discussion activities and forecasting future activities. Comparability means that a learner can compare his/her relative position with other learners' performance levels. Visualizations of online discussion activity results, based on learning analytics, areprovided in the form of a dashboard. As such, it is necessary to supply a large amount of information in a limited space. Implicity refers to elimination of unnecessary elements from the physical, visual, and cognitive aspects of information, and expressing meaningful information in abbreviated forms (Lohr, 2007). Overview plus detail involves providing detailed information as part of the discussion activity's full overview (Shneiderman, 1996).

In this study, these visual design guidelines used in previous studies to develop learning analytics dashboards were analyzed according to visual feedback information (see Table 3). Our dashboards were then developed in accordance with these guidelines.

Table 3. Visual design guidelines according to visual feedback information types

	Traceability	Comparability	Implicity	Overview+Details
Participation	Visualize learner participation levels according to the allocated discussion time period for a particular topic (Bakharia et al., 2016)	Provide the average and highest levels of participation to allow learners to compare their participation with their peers (Beheshitha, Hatala, Gašević, & Joksimović, 2016)	Visualize participation levels using color symbols (e.g., green for good, yellow for fair, and red for poor) (Wise et al., 2013)	Provide levels of team participation and all learner participation, and make the details visible by selection (Erickson & Kellogg, 2003)
Interaction	Visualize interaction levels between learners over time (Schneider, Passant, & Decker, 2012)	Be able to compare the level of interaction between learners (Mochizuki et al., 2007)	Visualize interaction levels using visual elements, such as location, size, color, and brightness (Hara et al., 2000)	Use a sociogram to allow for comparisons of interaction patterns within and among teams (Erickson & Kellogg, 2003)

Keywords	Make it possible to see changes in frequently mentioned keywords over time (Yi, Kang, Stasko, & Jacko, 2007)	Present the extent to which keywords are mentioned in the discussion to allow for comparisons with other learners (Mochizuki et al., 2007)	Simplify the central word using the word cloud technique (Siemens & Baker, 2012)	Visualize the degree to which students mention keywords (Teplovs, 2008) and the overall distribution of keywords in the discussion
Message types	Visualize the distribution of message types by discussion topics (Bakharia et al., 2016)	Present a distribution of message types that can be compared to the average of other learners (Tan et al., 2017)	Visualize the distribution of message types in the form of a radial graph (Ferguson, 2012)	Provide message types for all learners, and if one has selected each type, make related opinions available for viewing
Pros/cons message types	Visualize the distribution of pros and con opinions over time	Make it easy to compare the pro and con opinions on the discussion topic (Teplov, 2008)	Mark the pros and cons with different symbols or colors	Provide message types (pro and con) for all learners, and if one has selected each type, make related opinions available for viewing

Based on the above guidelines, the specific research questions were as follows:

- What are the appropriate learning analytics dashboards for online discussion activities that correspond with the aforementioned visual feedback information?
- How do learners and instructors perceive online discussion dashboards?

3. Research method

Lantz (1985) defined prototyping methodology as a "system development methodology based on building and using a model of a system for designing, implementing, testing and installing the system" (p. 1). In this methodology, after a succinct statement of objectives and goals, development is conducted using parallel processes through which prototype designs are created. The prototyping process requires having the system's definitions, an opportunity to use and test the prototype, and software that allows the rapid building and modification of the prototype (Tripp & Bichelmeyer, 1990). Often, the initial prototype emphasizes only the visual aspects of the final product, because these are less costly and demanding to build. After overall format decisions have been made, an executable prototype may be constructed to determine the product's usability (Jones & Richey, 2000). This process can be used to verify a product's form, fit, and function. It also has a strong impact on productivity, that is, means getting a product from concept to prototype to reality (Kamrani & Nasr, 2010). Following this methodology, we designed, developed, reviewed, and revised several learning analytics dashboards in parallel, using an iterative process. Figure 1 shows the specific sequence used in this study for each dashboard.

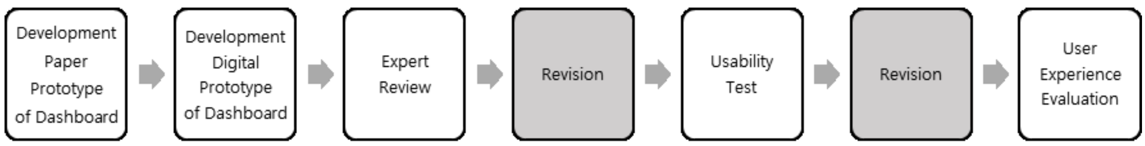


Figure 1. The study procedures

4. Initial visual learning analytics dashboards for online discussions

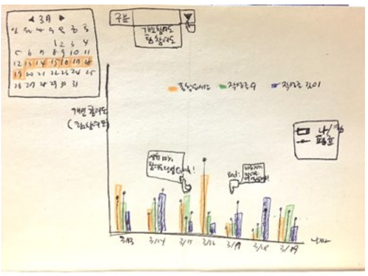
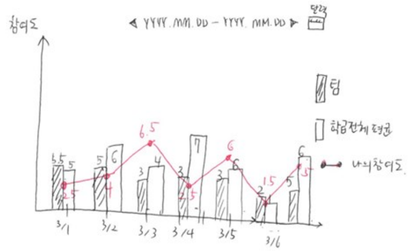
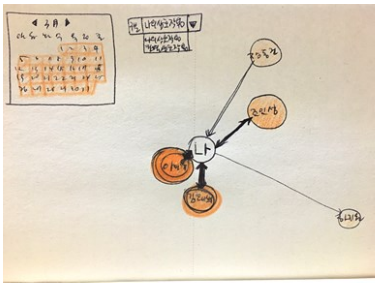
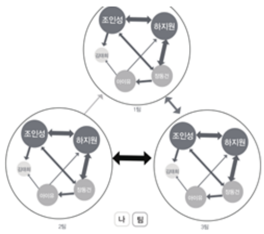
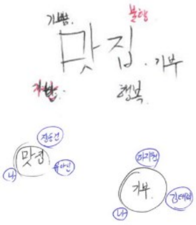
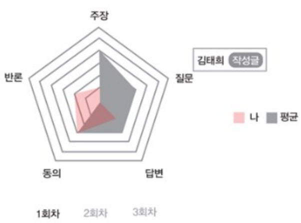
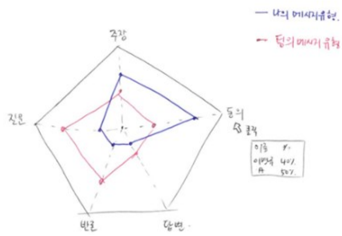
4.1. Participants and procedures

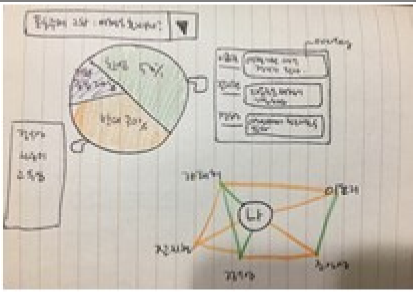
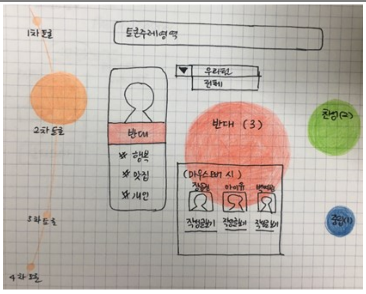
Six designers were invited to design paper prototypes for the learning analytics dashboards. Paper prototyping is widely used for designing, testing, and refining user interfaces, allowing designers to stay focused on users, while requiring little or no programming skills on the part of the designers (Rudd, Stern, & Isensee, 1996). The participants were three instructional designers, two web designers, and a computer programmer. They were provided with a fictional scenario and design guidelines, including the goal of the dashboards for online discussion activities and design specifications in an online format. The design process started with a request to develop these dashboards, and it then defined the target users, presented their persona, and offered examples of how to use the dashboards. Each designer created five types of learning analytics dashboards: participation, interaction, keywords, discussion message types, and debate (pros/cons) message types.

4.2. Paper prototypes

Although the designers were asked to develop paper prototypes, two used a digital form due to the convenience of computer authoring tools. Table 4 shows some of the final paper prototypes, chosen after the expert reviewing stage.

Table 4. A selection of paper prototypes developed by the six designers

Visual information	Examples of paper prototypes	
Participation		
Interaction		 
Keywords		
Message types		

Pros/cons message types		

4.3. First digital prototypes

A learning analytics dashboard, the most widely used intervention strategy in learning analytics, is a visual display that provides information presenting students' learning processes and behavior patterns (Jo, 2012). We designed and developed the first digital prototypes of the learning analytics dashboards to support individuals' self-reflective learning based on case studies (Yoo & Jin, 2017), theoretical reviews on self-regulating in the learning analytics approach (Sung, Jin, & Yoo, 2016), and learning theories, such as cognitive theory, cognitive-behaviorism, and social constructivism. The visual guidelines were applied to the five learning analytics dashboards, designed as follows (Figure 2).

The *participation dashboard* represents learners' participation in a particular discussion topic alongside the average participation levels of the team and the entire class. Based on Bandura (1991), it enables self-evaluation through intrinsic reinforcement, as self-regulation denotes the ways in which an individual may influence their external environment through self-observation and self-judgment. By applying the traceability guideline, when a specific period is selected, the participation level in that period is displayed as a bar graph, allowing learners to see their participation level on a specific date. After comparability implementation, a learner could compare the participation levels of all classmates, the team, and him/herself (traffic-light colored circles and line graph). The individual participation value is displayed in each circle: green means excellent, yellow is good, and red is a warning.

The *interaction dashboard* allows learners to identify the interaction levels among learners and teams by discussion topic. It represents connectivism in that it depends on building networks of information and supports a social constructivist pedagogy by proposing social discussion forums where learners can connect with others and exchange information (Khalil & Ebner, 2016). Here, the traceability guideline assesses and displays interaction patterns over a defined period, and the comparability guideline allows for interaction comparisons between the team and the whole class. The implicity guideline was followed by displaying interaction level differences through color, brightness, and size (a longer and thicker line corresponds to a higher interaction level). Learners can understand inter-team and interpersonal interactions through the overview plus detail guideline.

The keywords and message types dashboards are informed by the cognitive-behaviorist model, which involves providing students with guided learning and feedback (Khalil & Ebner, 2016). Through the *keywords dashboard*, learners can not only identify frequently used keywords over a certain period, but also discover who has used them often and compare them with their own frequently used terms. By applying the four visual design guidelines, the keywords can be identified for each discussion topic and the degree to which a keyword is mentioned can be compared using circle sizes. These circles are located near the keywords and appear larger as the keywords are mentioned more frequently. The green circle represents "Me" and the colors (green, orange, or blue) indicate the different teams.

The *message types dashboard* helps learners identify the type of messages used in discussion topics. Previous research has distinguished different message types by analyzing the content of online discussions. Many researchers have drawn on Henri's (1992) discussion analytical model, using message analysis frameworks based on the type of discussion. The analytical model used in this study followed the recommendations of Cho, Park, Kim, Suk, and Lee (2015) to divide message types into five categories: statement, agreement, argument, question, and answer. To facilitate message type comparisons, their distribution is visualized as a radial graph. "My message type" is represented in orange, and team message types are in blue. When each message type is clicked, learners can see a list of relevant posts and read each post.

The *pros and cons dashboard* helps the learners see how many people are on each side of a debate-style forum discussion, who wrote pros and cons, and what they wrote. Previous studies revealed that providing a pros and cons dashboard has a positive effect on satisfaction in discussion activities, according to students' writing efficacy and social comparative motivations (Jin & Yoo, 2019). In this dashboard, learners can see the number distributions of the pros and cons, and the ratio by debate topic. The pros were represented as blue and the cons as red. When learners click "pros" or "cons," they can view a list of learner names that have given their opinion and read them.

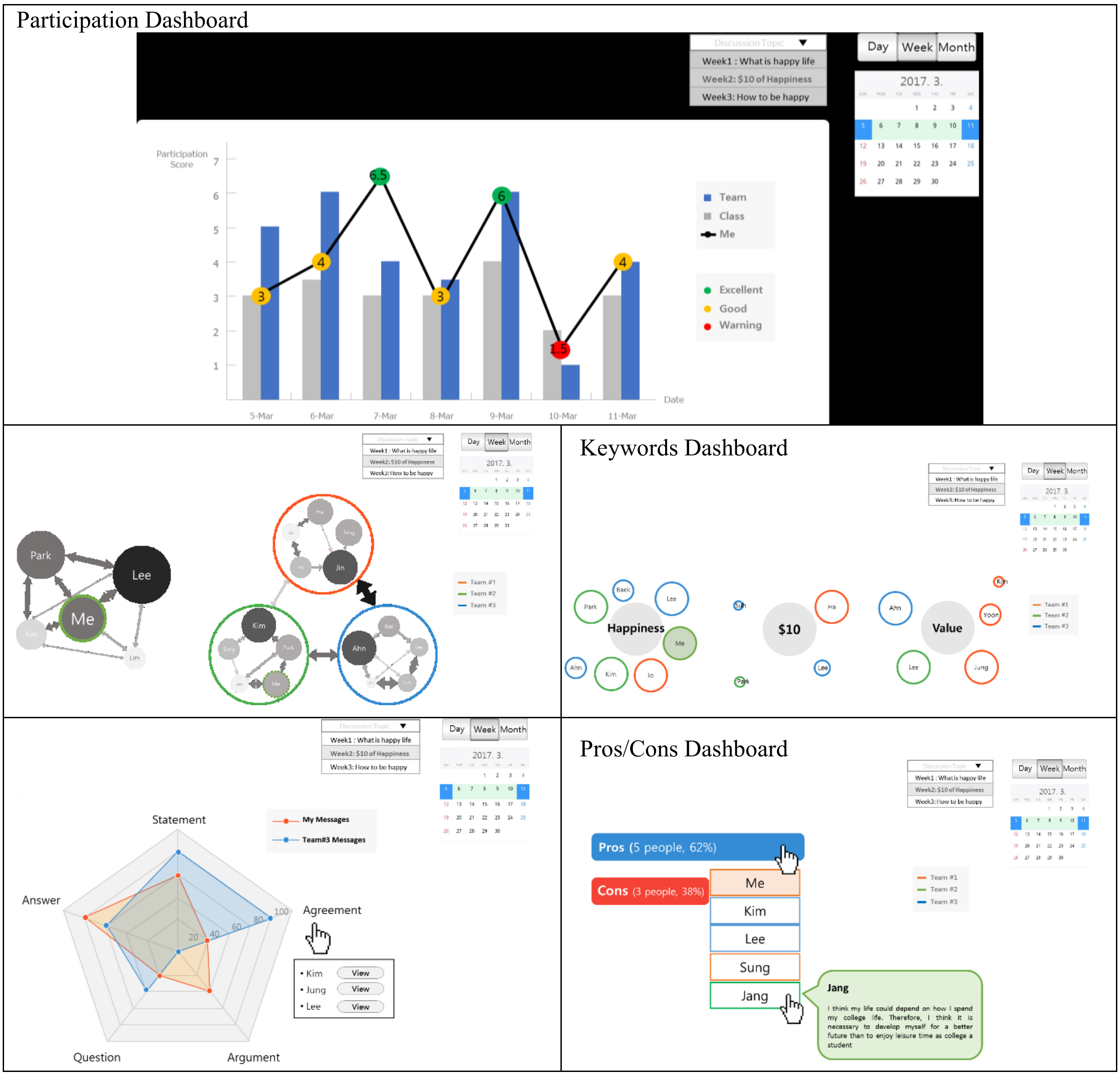


Figure 2. First version of digital prototypes

5. Revision and validations of learning analytics dashboards

5.1. Expert review participants and procedure

The expert review was conducted by six experts, who were recruited for their theoretical and practical experience in educational technology or learning analytics. The experts were introduced to the purpose of this research, the visual design guidelines, and the learning analytics dashboards. We conducted semi-structured interviews to determine which was the most incomprehensible prototype, and what were the strengths and weaknesses of each. We also asked participants their opinions and suggestions to improve the dashboards. Their reviews and comments were collected and analyzed for each dashboard type.

5.2. Expert review results

The experts' comments on each learning analytics dashboard are presented in Table 5. We reflected on the expert reviews and assessed the learning theories and visual design guidelines of previous studies to revise the learning analytics dashboards.

Table 5. Expert review results

Category	Answers
Participation dashboard	The line graph in the participation dashboard seems to indicate the changes in participation, but it would be better to delete it, because these changes are not the focus of this visualization (Expert A). The participant scores should not be directly presented as a specific number. Instead, they should be provided only when a learner wishes (Expert F).
Interaction dashboard	The interaction dashboard was not easy to understand. The interaction levels could be sufficiently expressed through the arrows' different thickness levels and colors (Expert C).
Keyword dashboard	The circles indicating "Me," displayed in green, were not visible and we recommend that they should be displayed in red for emphasis (Experts D & F).
Pros/Cons dashboard	The pros and cons distributions would be better represented as a pie chart in which the sum of the two (pros and cons) equals 100% (Expert F).

5.3. Participants and usability test procedure

We conducted a usability test with 20 graduate learners majoring in educational technology at university A and five instructors. Of the six experts who participated in the expert review, five also participated in the usability test, with the exception of Expert F. Usability is usually considered to be a user's ability to successfully carry out a task with a product, and the usability test serves to improve a tested product's use (Dumas & Redish, 1999). Our test consisted of five scales: accessibility (2 items; α = .85), usefulness (4 items; α = .95), satisfaction (2 items; α = .83), aesthetics (2 items; α = .73), and intention of use (2 items; α = .82) (Lund, 2001; Nokelainen, 2006). The Table 6 provides an example of a representative questionnaire.

Table 6. Usability test questionnaire

Category	Questionnaire
Accessibility	I can get the information what I want to know using the learning analytics dashboard.
Usefulness	It is easy to use.
Satisfaction	The information obtained from the learning analytics dashboard is valuable.
Aesthetic	I think the learning analytics dashboard is attractive.
Intention of use	I would recommend it to a friend.

Using a five-point Likert scale, the participants were asked to reflect on how they felt looking at the static dashboard image. Semi-structured interviews were also conducted to gather participants' additional comments on the dashboards. The interview questions, used if the previously determined answers were inconclusive, were based on the following three questions: which dashboard was the most understandable, which was the most incomprehensible, and what were the strengths and weaknesses of each dashboard? In this process, we employed a stimulated recall interview, which is useful in helping users recall specific moments during the test (Park & Jo, 2015).

5.4. Usability test results

Table 7 shows the results of the usability test. The users' perceptions of the dashboards were generally positive. Both learners and instructors responded that the learning analytics dashboards were useful for obtaining information, easy to use, save time, and that they would recommend them to friends. Table 8 presents the interview results.

After the instructor and learner interviews, expert reviews were performed again. Expert B recommended removing the date options, because most discussion activities are provided to learners by topic rather than by date. Experts C and E suggested fixing the positions of the circle representing each team in the keywords dashboard, because it was difficult to recognize at a glance where each team is positioned.

Table 7. Usability test results

Category	Students		Instructors	
	M	*SD*	*M*	*SD*
Accessibility	4.23	0.77	4.50	0.53
Usefulness	4.45	0.63	4.45	0.52
Satisfaction	4.00	0.72	4.40	0.70
Aesthetics	4.23	0.53	4.50	0.53
Intention of use	4.25	0.71	4.50	0.53

Table 8. Usability interview results

Category	Answers
Participation dashboard	• It seems unnecessary to express individual participation with a line graph, because it is already shown as a circle (Instructor B).
Interaction dashboard	• The fact that information on the interactions of the whole class and the team is provided simultaneously is confusing (Instructor A). • Understanding what the visualization elements meant was time-consuming. I was confused by the simultaneous appearance of interpersonal and team interactions on the same screen (Student A). • It is difficult to understand what the size of the circle representing an individual meant. If it represents participation, it seems to conflict with the interaction (Instructor C). • Too many colors were applied to the dashboard, making it difficult to grasp at a glance what each color represented (Student C).
Keyword dashboard	• The interaction dashboard provides information for both teams and individuals, but the keyword dashboard provides only one. I would like these to be separate here as well (Student A). • It was very complicated to visualize the extent of all learners' messages in the class (Student B).

5.5. Final version of the learning analytics dashboards

At this stage, the learning analytics dashboards were revised to reflect the research participants' opinions, gathered through expert reviews and usability tests. Figures 3, 4, 5, 6, and 7 presents the final revised versions.

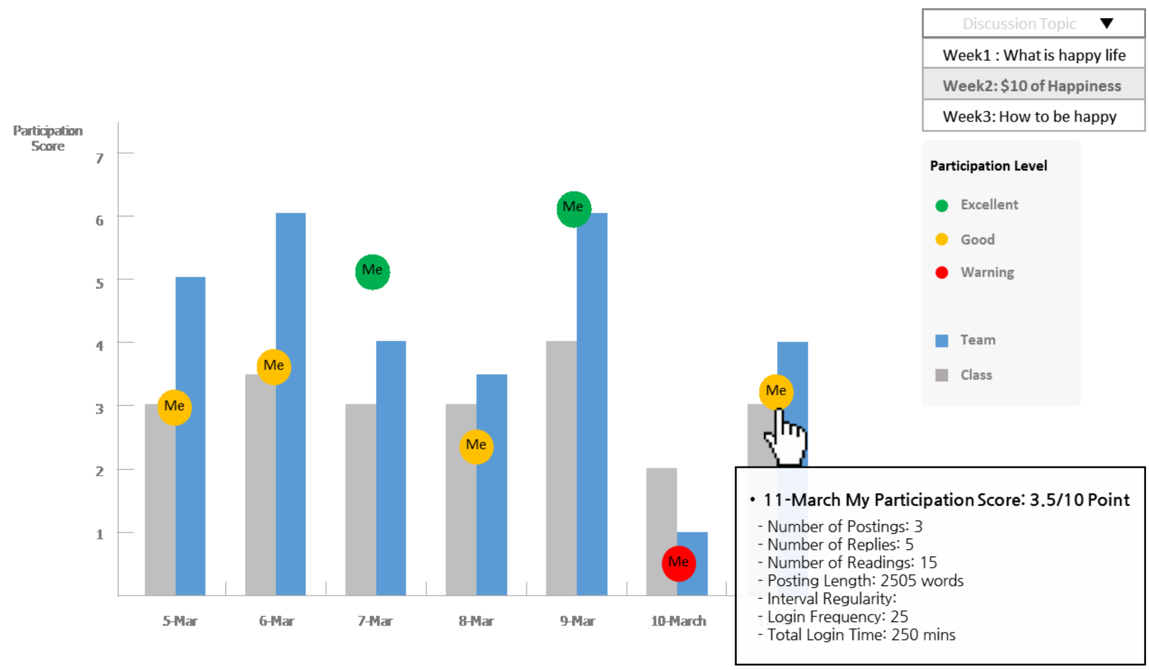


Figure 3. Final revised version of the Participation Dashboard

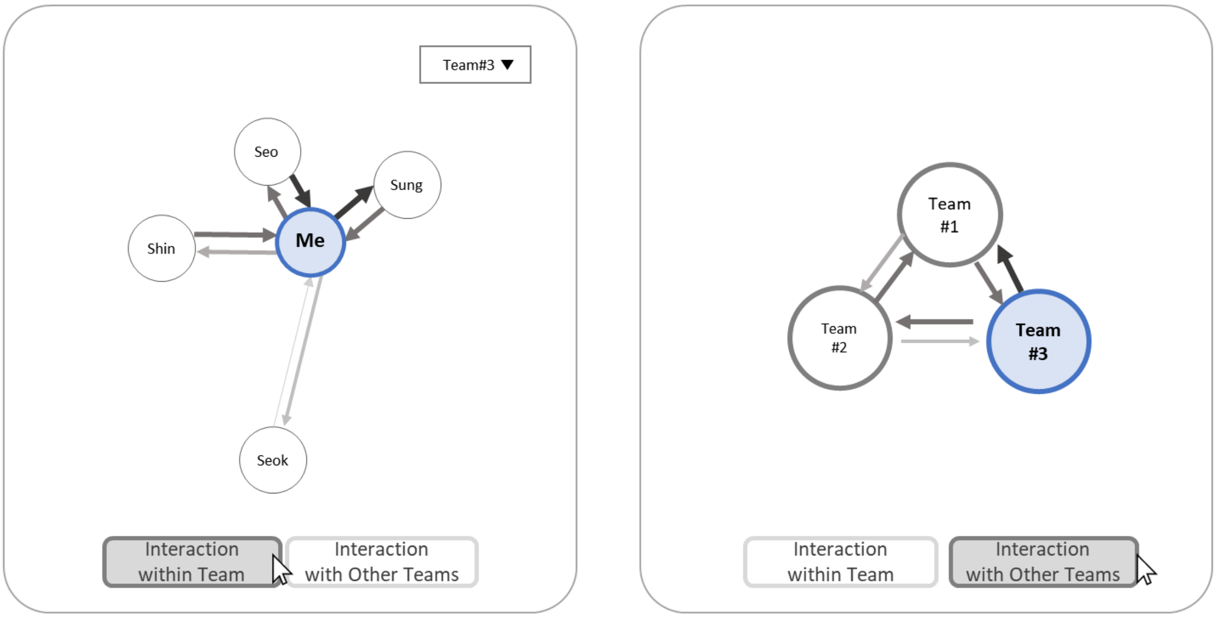


Figure 4. Final revised version of the Interaction Dashboard

The following discussion enumerates the significant changes we made from the first version of the learning analytics dashboards.

First, we applied the traceability principle and provided two options for selecting discussion topics and data, which allowed the learners to reflect on their previous discussion activities. We deleted the date selection, retaining only the discussion topic selection option, because too many options can confuse the users.

For the participation dashboard, we deleted the line graph because instructor B indicated that individual participation is already shown as a circle (Figure 3). The individual participation score was removed and it presented as a separate pop-up window that appears when a user hovers the mouse cursor over the circle. When a learner hovers over the graph, the score displays on the graph.

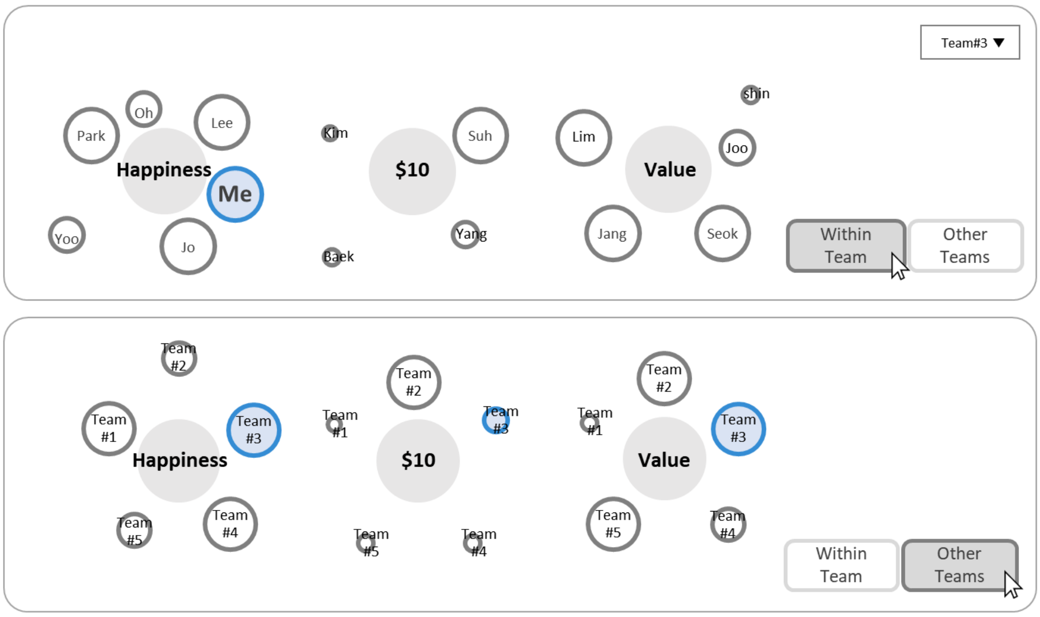


Figure 5. Final revised version of the Keywords Dashboard

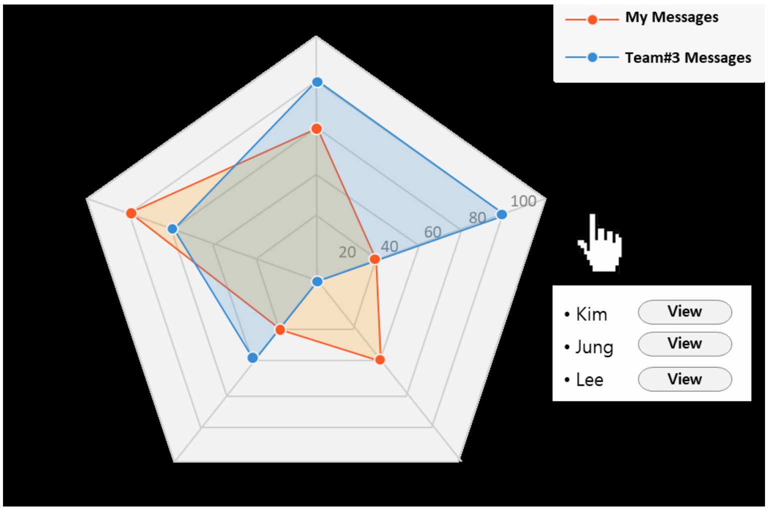


Figure 6. Final revised version of the Message Types Dashboard

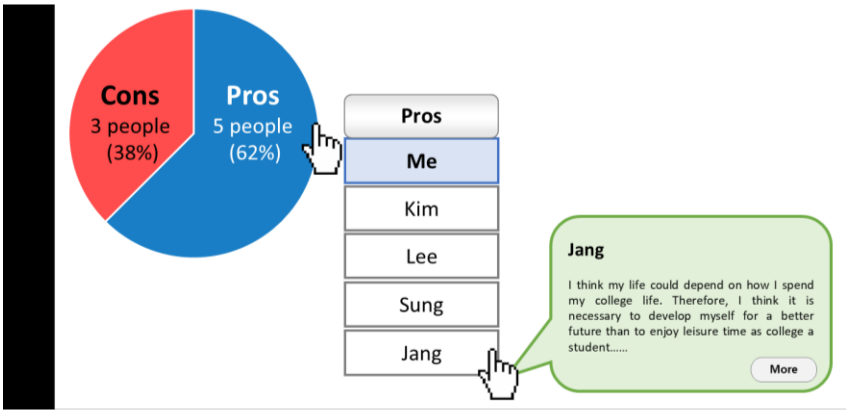


Figure 7. Final revised version of Pros/Cons Dashboard

In the first version of the interaction dashboard, different colors were used for each team, but these were simplified and only the colors of "my team" were highlighted in the final version (Figure 4). The interaction level was represented by the thickness and color of the arrows, and the interactions between teams were depicted more simply. In addition, the "interaction within team" and the "interaction with other teams" screens were divided so that each would be displayed only when the learner selects them, since it appeared to be confusing to have both visualizations on one screen.

Keyword frequency was visualized within and across teams, and a team's spatial position was fixed in order to maintain the integrity of the interaction dashboard (Figure 5). The message types dashboard helps learners identify the type of messages used in discussion topics. To facilitate message type comparisons, a radical graph shows the distributions of message types (Figure 6). For pros/cons dashboard, the bar graph was changed to a chart, which made it easier to see what percentage of pros and cons constituted the total 100% of the opinions (Figure 7).

We conducted expert validations to ensure the learning analytics dashboards' reliability. The experts were presented with the final versions of the dashboards, and were asked to evaluate whether the instructors' and learners' feedback was well-applied to the final version. The expert validation results were calculated using a content validity index (Chang, Gardner, Duffield, & Ramis, 2010). The agreement for all items was found to be 100%. Hence, each dashboard was found to be suitable for achieving the respective design purpose.

6. User experience evaluations of the learning analytics dashboards

6.1. Participants and procedures

After the learning analytics dashboards were created, we conducted a user experience evaluation that would reflects a broader perspective, assessing the individual's entire interaction with a product as well as his/her thoughts, feelings, and perceptions (Albert & Tullis, 2013). The participants comprised 31 graduate learners majoring in educational technology at university B and five instructors who participated in the usability test. We used the user experience questionnaires and data analysis tool developed by Laugwitz, Held, and Schrepp (2008; www.ueq-online.org) and included 26 items pertaining to the following scales: attractiveness (6 items; α = .86), perspicuity (4 items; α = .70), efficiency (4 items; α = .57), dependability (4 items; α = .77), stimulation (4 items; α = .84), and novelty (4 items; α = .63). The items are scaled from -3, the most negative answer, to +3, the most positive answer. This study sought to evaluate the relative quality of a user's experience through a benchmark analysis.

6.2. User experience evaluation results

A benchmarking analysis was conducted to compare these results with the user experience results for other products. Figure 8 shows the results of a benchmarking analysis of the learning analytics dashboards, both learners and instructors had a mostly positive impression of the dashboards for all the categories. The learners' overall impressions of the final versions were almost average. The novelty was excellent, efficiency and stimulation were good, perspicuity was above average, and attractiveness and dependability were below average. The instructors' impressions were excellent except for attractiveness. Both the instructors and the learners recognized the learning analytics dashboards as being very creative and innovative (novelty). They also confirmed that the dashboards were efficient in helping them understand the discussion activities and that they were motivated to use them (efficacy and simulation). In addition, their level of familiarity with the dashboards was confirmed to be above average (perspicuity). As their perceptions of dependability were relatively low, there is room for improvements that would allow users to have more control over their interactions with the dashboards (dependability).

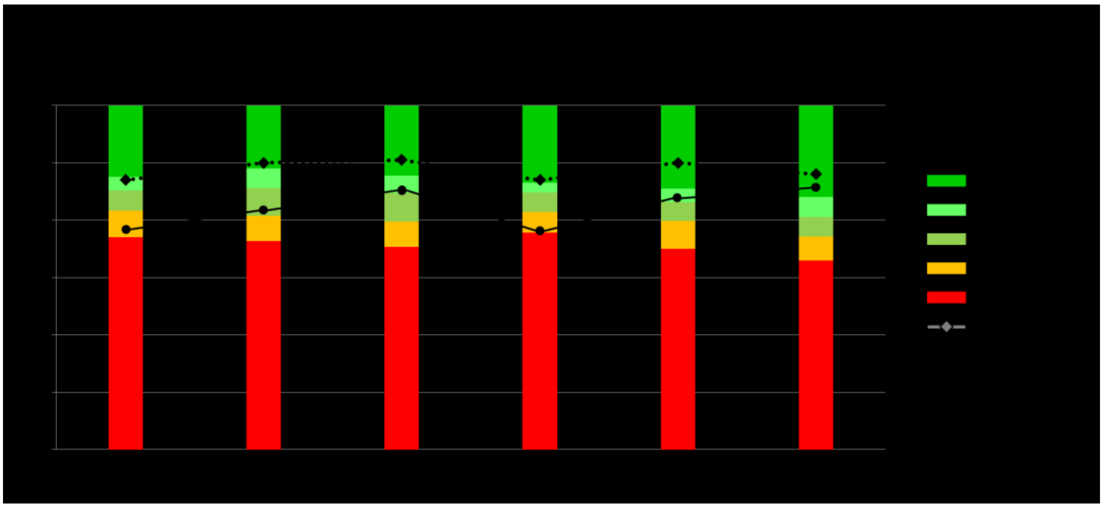


Figure 8. Benchmarking analysis results

Table 9. Results of user experience interviews

Category	Answers
Familiarity	The participation dashboard was the most familiar, probably because the bar graph is a familiar to me (Student A).
Participation engagement	The participation dashboard helps me understand the participation levels of an entire class, a team, and my own. I think it can be a stimulus for the next discussion and I would like to use it again (Student D).
Quality of discussion	I worry about the discussion message's quality if unnecessary or meaningless dialogue is included at the participation level (Instructor D).

Regarding the user experience interviews, we posed semi-structured interview questions that included participants' previous experience using learning analytics dashboards and their feelings about the experience of each dashboard. Table 9 presents a summary of the user experience interviews.

7. Discussion

7.1. Contributions to practice

The learning analytics dashboards developed in this study can serve as a guide regarding *what* to visualize in online discussion activities and *how* to do so. Yau (2013) argued that one must determine what users want to know before deciding which visualization technique to use, in order to take advantage of the appropriate visualization techniques for the required information. In other words, to develop an effective dashboard for online discussion activities, it is important to identity what information learners and instructors want to know. This study identified the following five kinds of information as useful in online discussion environments: participation, interaction, keywords, discussion message types, and distributions of pros and cons in a debate format. Our development of visual design guidelines and five types of dashboards has evident practical implications for future researchers interested in developing learning analytics dashboards for discussion activities.

The dashboards developed in this study presents learners with qualitative information on discussion contents and quantitative participation information regarding online discussion activitie. Furthermore, the content on learning activities can be classified into quantitative information on "how much is learned" and qualitative information on "what to learn" (Mayer, 2011). Previous research on dashboard development has mostly visualized quantitative information by analyzing learning log data. We placed a particular emphasis on the importance of qualitative analysis, because fragmented information, such as the number of postings and responses in online discussion activities, does not provide sufficient information for learners to reflect on their discussion activities (Malheiro, Morgado, & Mendes, 2008).

Recent research has found that applying learning analytics to learning activities can improve students' levels of engagement, which can in turn play an essential role in a self-regulated learning environments (Lai & Hwang, 2016). The learning analytics dashboards presented in this study can provide learners with feedback on their cognitive and social engagement in online discussion activities. Cognitive engagement refers to "the learner's psychological investment in and effort directed toward learning, understanding, or mastering the knowledge, skills, or crafts that academic work is intended to promote" (Newman, Wehlage, & Lamborn, 1992, p. 12). Social engagement denotes the process of communicating within a discussion environment in an online community. It is expected that, by providing learners with social engagement information, the learning analytics dashboards will promote a sense of cohesion and group-belonging in learners during online discussion activities, thereby alleviating feelings of isolation or alienation in online learning environments.

7.2 Empirical contributions

This study has implications for identifying the cognitive, emotional, and social problems that learners and instructors experience during online discussion activities. We included a user experience evaluation of the learning analytics dashboards to describe the interactions between a user and a product (Albert & Tullis, 2013). The results of this evaluation show lower scores for learners than for instructors. It was confirmed that most learners were using these kinds of learning analytics dashboards for the first time. Therefore, they may have felt that most dashboards were useful, but a few (e.g., the interaction dashboard) may have been unfamiliar or relatively less dependable. By contrast, the instructors, whose prior experience in online discussion activities was brought to bear on the operation and evaluation of the learning analytics dashboards, made very positive assessments of all the scales (perspicuity, efficiency, dependability, stimulation, and novelty). This result suggests an improvement plan from the viewpoint of various users, and indicates that learners should be informed about the use of prototypes in educational environments.

Schrepp, Hinderks, and Thomaschewski (2014) provided a benchmarking data set for user experience evaluations of various new systems or services, and a basis for analyzing the evaluation results. However, their benchmarking dataset did not include the user experience evaluation results of dashboards based on learning analytics. From this point of view, the data collected in this study can serve as a base that can be used to compare user experience evaluations in future studies that develop learning analytics dashboards.

7.3 Limitations and future research

Despite the extensive processes of this research, the study has certain limitations. First, it is necessary to implement the learning analytics dashboards developed in this study within an online learning system and investigate their educational effects. This study provides conceptual modeling and visualization techniques according to five types of visualization objects on discussion activities. In order to implement this in e-learning systems, computational modeling should be done using learning log data, learner creation data, and instructor creation data from e-learning environments. Further research should be conducted to analyze whether the learning analytics dashboards implemented in an online discussion learning system are effective in achieving educational objectives. In the future, it will be useful to apply quantitative evaluation methods, such as eye tracking and physiological response measurement. Second, only 20 learners from a specific university class and five instructors participated in the usability test in this study, which limits the generalizability of our usability test results. Finally, this study did not consider the possibility of providing interventions for learners' individual differences. As such, for the next phase of designing and developing learning analytics dashboards, individual differences in achievement levels need to be considered (Park & Jo, 2015).

Acknowledgement

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2015S1A5A2A03048269, NRF-2019S1A5A2A03053308).

References

Albert, W., & Tullis, T. (2013). *Measuring the user experience: Collecting, analyzing, and presenting usability metrics*. Waltham, MA: Elsevier Inc.

Ali, L., Hatala, M., Gašević, D., & Jovanović, J. (2012). A Qualitative evaluation of evolution of a learning analytics tool. *Computers & Education*, *58*(1), 470-489.

Alabi, H., Code, J., & Irvine, V. (2013). Visualizing learning analytics: Designing a roadmap for success. In *Proceedings of EdMedia+ Innovate Learning* (pp. 951-959). Waynesville, NC.

An, H., Shin, S., & Lim, K. (2009). The Effects of different instructor facilitation approaches on students' interactions during asynchronous online discussions. *Computers & Education*, *53*(3), 749-760.

Bakharia, A., Corrin, L., de Barba, P., Kennedy, G., Gašević, D., Mulder, R., Williams, D., Dawson, S., & Lockyer, L. (2016). A Conceptual framework linking learning design with learning analytics. In *Proceedings of the 6th International Learning Analytics & Knowledge Conference* (pp. 329-338). New York, NY: ACM.

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes, 50*(2), 248-287.

Beheshitha, S. S., Hatala, M., Gašević, D., & Joksimović, S. (2016). The Role of achievement goal orientations when studying effect of learning analytics visualizations. In *Proceedings of the 6th International Learning Analytics & Knowledge Conference* (pp. 54-63). New York, NY: ACM.

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

Buckley, F. (2011). Online discussion forums. *European Political Science, 10*(3), 402-415.

Chang, A. M., Gardner, G. E., Duffield, C., & Ramis, M. A. (2010). A Delphi study to validate an advanced practice nursing tool. *Journal of Advanced Nursing, 66*(10), 2320-2330.

Cho, Y., Park, H., Kim, J., Suk, Y., & Lee, S. (2015). Exploring roles of feedback to facilitate online discussion. *Asian Journal of Education, 16*(2), 289-313.

Dawson, S., Bakharia, A., & Heathcote, E. (2010). SNAPP: Realising the affordances of real-time SNA within networked learning environments. In *Proceedings of the Seventh International Conference on Networked Learning 2010* (pp. 125-133). Lancaster, UK: Lancaster University.

Dumas, J. S., & Redish, J. (1999). *A Practical guide to usability testing.* Portland, OR: Intellect books.

Erickson, T., & Kellogg, W. A. (2003). Social translucence: Using minimalist visualisations of social activity to support collective interaction. In K. Höök, D. Benyon, & A. J. Munro (Eds.), *Designing Information Spaces: The Social Navigation Approach* (pp. 17-41). London, UK: Springer.

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

Govaerts, S., Verbert, K., Duval, E., & Pardo, A. (2012). The Student activity meter for awareness and self-reflection. In *Proceedings of CHI'12 Extended Abstracts on Human Factors in Computing Systems* (pp. 869-884). New York, NY: ACM.

Hara, N., Bonk, C. J., & Angeli, C. (2000). Content analysis of online discussion in an applied educational psychology course. *Instructional Science*, *28*(2), 115-152.

Harasim, L. (2000). Shift happens: Online education as a new paradigm in learning. *The Internet and Higher Education, 3*(1-2), 41-61.

Harasim, L. (1993). *Global networks: Computers and international communication*. Cambridge, MA: MIT Press.

Hatala, M., Beheshitha, S. S., & Gasevic, D. (2016). Associations between students' approaches to learning and learning analytics visualizations. In *Proceedings of the LAK 2016 Workshop on Learning Analytics for Learners* (pp. 3-10). Aachen, Germany: RWTH Aachen University.

Henri, F. (1992). Computer conferencing and content analysis. In A. R. Kaye (Ed.), *Collaborative Learning through Computer Conferencing* (pp. 117-136). Berlin, Germany: Springer.

Hew, K. F., & Cheung, W. S. (2011). Higher-level knowledge construction in asynchronous online discussions: An Analysis of group size, duration of online discussion, and student facilitation techniques. *Instructional Science, 39*(3), 303-319.

Ifenthaler, D. (2017). Are higher education institutions prepared for learning analytics? *TechTrends, 61*(4), 366-371.

Janssen, J., Erkens, G., Kanselaar, G., & Jaspers, J. (2007). Visualization of participation: Does it contribute to successful computer-supported collaborative learning? *Computers & Education*, *49*(4), 1037-1065.

Jin, S., Yoo, M., & Kim, T. (2015). Design of a dashboard for enhancing e-learning participation and interaction. *Journal of Educational Technology, 31*(2), 191-221.

Jin, S., & Yoo, M. (2019). The Effect of an online debate dashboard on students' satisfaction and perception. *Journal of Learner-Centered Curriculum and Instruction, 19*(7), 41-61.

Jo, I. (2012). On the LAPA (Learning Analytics for Prediction & Action) model suggested. In *Proceedings of the Future Research Seminar* (pp.1-6). Seoul, Korea: Korea Society of Knowledge Management.

Jones, T. S., & Richey, R. C. (2000). Rapid prototyping methodology in action: A Developmental study. *Educational Technology Research and Development, 48*(2), 63-80.

Kamrani, A. K., & Nasr, E. A. (2010). *Engineering design and rapid prototyping.* New York, NY: Springer.

Kehrwald, B. (2008). Understanding social presence in text-based online learning environments. *Distance Education, 29*(1), 89-106.

Khalil, M., & Ebner, M. (2016). What massive open online course (MOOC) stakeholders can learn from learning analytics? In M. Spector, B. Lockee, & M. Childress (Eds.), *Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy* (pp. 1-30). Cham, Switzerland: Springer.

Lai, C. L., & Hwang, G. J. (2016). A Self-regulated flipped classroom approach to improving students' learning performance in a mathematics course. *Computers & Education, 100*, 126-140.

Lambropoulos, N., Faulkner, X., & Culwin, F. (2012). Supporting social awareness in collaborative e-learning. *British Journal of Educational Technology, 43*(2), 295-306.

Lantz, K. E. (1985). *The Prototyping methodology.* Englewood Cliffs, NJ: Prentice Hall.

Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and Usability Engineering Group* (pp. 63-76). Berlin, Germany: Springer.

Lohr, L. L. (2007). *Creating graphics for learning and performance: Lessons in visual literacy.* Upper Saddle River, NJ: Prentice Hall.

Lund, A. M. (2001). Measuring usability with the use questionnaire. *Usability Interface, 8*(2), 3-6.

Malheiro, S., Morgado, L., & Mendes, A. (2008). Analysis of engaged online collaborative discourse: A Methodological approach. In A. J. Mendes, I. Pereira, & R. Costa (Eds.), *Computers and Education: Towards Educational Change and Innovation* (pp. 33-43). London, UK: Springer.

Mayer, R. E. (2011). *Applying the science of learning*. Boston, MA: Pearson/Allyn & Bacon.

Mochizuki, T., Kato, H., Fujitani, S., Yaegashi, K., Hisamatsu, S., Nagata, T., Nakahara, J., Nishimori, T., & Suzuki, M. (2007). Promotion of self-assessment for learners in online discussion using the visualization software. In N. Lambropoulos,

& P. Zaphiris (Eds.), *User-centered design of online learning communities* (pp. 365-397). Hershey, PA: Information Science Publishing.

Murray, T. (2014). Supporting deeper deliberative dialogue through awareness tools. In *Proceedings of the 2014 Build Peace Conference* (pp. 1-8). Cambridge, MA: MIT Media Lab.

Newman, E. M., Wehlage, G. G., & Lamborn, S. D. (1992). Taking students seriously. In F. E. Newmann (Ed.), *Student engagement and achievement in American secondary schools* (pp. 40-61). New York, NY: Teachers College Press.

Nokelainen, P. (2006). An Empirical assessment of pedagogical usability criteria for digital learning material with elementary school students. *Educational Technology & Society, 9*(2), 178-197.

Pallotta, V., & Delmonte, R. (2011). Automatic argumentative analysis for interaction mining. *Argument & Computation, 2*(2-3), 77-106.

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

Rapport, M. (1991). *Computer mediated communication: Bulletin boards, computer conferencing, electronic mail, information retrieval.* San Francisco, CA: Wiley & Sons.

Rudd, J., Stern, K., & Isensee, S. (1996). Low vs. high-fidelity prototyping debate. *Interactions*, *3*(1), 76-85.

Sack, W. (2000). Conversation map: An Interface for very large-scale conversations. *Journal of Management Information Systems, 17*(3), 73-92.

Schneider, J., Passant, A., & Decker, S. (2012). Deletion discussions in Wikipedia: Decision factors and outcomes. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration* (pp. 1-10). New York, NY: ACM.

Schrepp, M., Hinderks, A., & Thomaschewski, J. (2014). Applying the user experience questionnaire (UEQ) in different evaluation scenarios. In *International Conference of Design, User Experience, and Usability* (pp. 383-392). Cham, Switzerland: Springer.

Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S., & Kirschner, P. A. (2020). Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior, 107,* 105512. doi:10.1016/j.chb.2018.05.004

Shneiderman, B. (1996). The Eyes have it: A Task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages* (pp. 336-343). IEEE. doi:10.1109/VL.1996.545307

Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In *Proceedings of the 2nd International Learning Analytics & Knowledge Conference* (pp. 252-254). New York, NY: ACM.

Sung, E., Jin, S., & Yoo, M. (2016). Exploring learning data for supporting self-directed learning in the perspective of learning analytics. *Journal of Educational Technology, 22*(3), 487-533.

Tan, J. P. L., Koh, E., Jonathan, C. R., & Yang, S. (2017). Learner dashboards a double-edged sword? Students' sense-making of a collaborative critical reading and learning analytics environment for fostering 21st century literacies. *Journal of Learning Analytics*, *4*(1), 117-140.

Teplovs, C. (2008). The Knowledge space visualizer: A Tool for visualizing online discourse. In *Proceedings of the Eighth International Conference of the Learning Sciences* (pp. 1-12). Utrecht, Netherlands: International Society of the Learning Sciences.

Tripp, S. D., & Bichelmeyer, B. (1990). Rapid prototyping: An Alternative instructional design strategy. *Educational Technology Research & Development, 38*(1), 31-44.

Verbert, K., De Croon, R., De Laet, T., Broos, T., Millecamp, M., Ochoa, X., Bodily, R., Kay, J., Drachsler, H., & Conati, C. (2019). VISLA: Visual approaches to learning analytics. In *Proceedings of the 9th International Learning Analytics & Knowledge Conference* (pp. 750-753). SoLAR.

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

Vieira, C., Parsons, P., & Byrd, V. (2018). Visual learning analytics of educational data: A Systematic literature review and research agenda. *Computers & Education, 122*, 119-135.

Weinberger, A., & Fischer, F. (2006). A Framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*(1), 71-95.

Wise, A. F., Zhao, Y., & Hausknecht, S. N. (2013). Learning analytics for online discussions: A Pedagogical model for intervention with embedded and extracted analytics. In *Proceedings of the 3rd International Learning Analytics & Knowledge Conference* (pp. 48-56). New York, NY: ACM.

Yau, N. (2013). *Data points: Visualization that means something.* Somerset, NJ: John Wiley & Sons.

Yi, J. S., Kang, Y., Stasko, J. T., & Jacko, J. A. (2007). Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics, 13*(6), 1224-1231.

Yoo, M., & Jin, S. (2017). An Analytic review of dashboard based on learning analysis for online discussion activities. *The Journal of Educational Information and Media, 23*(3), 465-497.

Daradoumis, T., & Arguedas, M. (2020). Cultivating Students' Reflective Learning in Metacognitive Activities through an Affective Pedagogical Agent. *Educational Technology & Society*, *23* (2), 19–31.

Cultivating Students' Reflective Learning in Metacognitive Activities through an Affective Pedagogical Agent

Thanasis Daradoumis[1,2*] and Marta Arguedas[1]
[1]Faculty of Computer Science, Multimedia and Telecommunications, Open University of Catalonia, Spain //
[2]Department of Cultural Technology and Communication, University of Aegean, University Hill, Mytilini, Greece // adaradoumis@uoc.edu // martaarg@uoc.edu
[*]Corresponding author

(Submitted April 2, 2020; Revised April 27, 2020; Accepted May 15, 2020)

ABSTRACT: There is an increasing interest in the ways pedagogical agents can provide cognitive, emotional, and metacognitive support to students. Moreover, several research studies have proposed various approaches for cultivating students' reflective learning. A variety of research has also been conducted into interrelations between metacognition and affective processes. However, very few studies have examined the effect of emotional feedback provided by a virtual Affective Pedagogical Tutor (APT) on students' self-reflection during a metacognitive learning activity. An experimental research design was used in a study aimed at measuring the extent to which an APT's affective feedback managed to enhance students' self-reflection about what and how they had learned. Participants were a sample of 45 fourth-year high school students, who were divided into experimental and control groups (APT vs. human tutor) in a real online learning situation that involved metacognitive activities. A questionnaire was specifically designed to collect data from both groups. Our results showed that experimental group students achieved better scores in the self-reflection process, since the APT's affective feedback significantly enhanced students' conceptual change (what has changed with respect to their initial beliefs), as well as students' personal growth and understanding (what led them to change their initial beliefs). They also indicated the affective competencies that the APT needs to have in order to achieve a conceptual and personal change in students. Finally, the limitations of our study and directions for future research are discussed.

Keywords: Reflective learning, Metacognitive activities, Affective tutor, Pedagogical agent, Affective feedback

1. Introduction

Reflective learning involves students' conscious thinking and analysis about what they have done in a previous learning activity (Henderson, Napan, & Monteiro, 2004). Reflective learning can be triggered and supported by specific metacognitive activities which allow students to become more engaged in their own learning (Carini, Kuh, & Klein, 2006). In particular, metacognitive activities enable students to enhance their awareness about the best practices they followed in order to learn more effectively as well as exercise essential skills such as critical and creative thinking, understanding and learning from failure, adaptability, personal responsibility, and more (Desautel, 2009; Schmitt, 1990). Consequently, we need to provide ways to help students engage in reflective learning in an efficient manner (Silver, 2013).

Several research studies have proposed approaches for cultivating students' reflective learning. One approach is students' explicit training through specific activities that aim to develop their metacognitive abilities (Jackson & Larkin, 2002; Lin, 2001; Parkes & Kajder, 2010). Another approach is based on social learning that encourages collaboration in small groups, which allows students to engage in a reflective practice that lets them comprehend their own learning in relation to others (Chinnery, Appleton, & Marlowe, 2019; Jarvela et al., 2015). Furthermore, a feedback approach can be used by the teacher, who can guide students through specific prompts to contemplate the learning process they have followed (Menz & Xin, 2016; Schoenfeld, 1992).

Emotions also play a very important role in motivation, self-regulated learning and performance (Arguedas, Daradoumis, & Xhafa, 2016a; Feidakis, Daradoumis, Caballé, Conesa, & Gañán, 2013; Pekrun, Goetz, Titz, & Perry, 2002). In metacognitive activities, a teacher's affective and cognitive feedback can make students reflect on the way they learn, the learning strategies they use and the way these strategies have influenced their learning. Moreover, in cases where students have had negative experiences, a teacher's affective feedback should help students progressively attenuate the impact that those negative experiences have had on their motivation for learning (Belland, Kim, & Hannafin, 2013).

In general, affective feedback should be timely, situation-aware and personal, so that it can redirect students' focus of attention and induce a change in the way they think, act and interact with others, while regulating their behavior in a learning situation (Bahreini, Nadolski, & Westera, 2012; Shen, Wang, & Shen, 2009). Moreover, adequate affective feedback depends on the teacher's emotional competencies (Jennings, 2011). The teacher can also provide affective feedback based on mechanisms that provide emotion awareness information, which ensures students' emotional safety and their engagement or persistence in the learning experience (Feidakis, Caballé, Daradoumis, Gañán, & Conesa, 2014).

To assist the human teacher, in past studies animated pedagogical agents have been extensively used to provide customized feedback with the aim of improving both students' emotional states and learning performance (Atkinson, 2002; Elliott, Rickel, & Lester, 1999; Stone & Lester, 1996). However, more recent studies on pedagogical agents, outlined in the literature review work by Heidig and Clarebout (2011), question the efficiency of pedagogical agents, in the sense that feedback provided by these agents does not necessarily motivate, interest, or support students' learning better than other simpler teaching and learning artifacts. Yet, the research question set by Heidig and Clarebout, which explores both the conditions and design issues under which pedagogical agents facilitate learner motivation and learning outcomes and when they are effective, still needs to be investigated further.

Consequently, the aim of this work is to measure the effectiveness of affective feedback types used by a human teacher and a virtual Affective Pedagogical Tutor (APT) in their group of students in activities promoting metacognitive (reflective) learning. We especially explore the effectiveness of the APT's affective feedback to identify what affective competencies the APT needs to have in order to effectuate a positive change on students' affective and cognitive states when performing metacognitive activities.

In order to achieve this goal, we have organized the rest of the paper as follows: First, we review the literature regarding affective pedagogical tutors, metacognition and affective feedback. Then, we set out our research context and questions. Next, we describe the methodology used in this study. Subsequently, we present the results and then discuss and analyze these results with respect to the research questions. Finally, we present the conclusions and future work.

2. Literature review on affective pedagogical tutors, metacognition and affective feedback

The area of pedagogical tutors has generated a significant amount of research, which has also proved to be quite controversial. Heidig and Clarebout's (2011) systematic review of pedagogical agents yielded no difference in learning. However, Schroeder, Adesope, and Gilbert's (2013) meta-analysis evidenced that pedagogical agents may produce a slight positive effect on learning performance. Furthermore, Schroeder, Romine, and Craig (2017) coincided with Heidig and Clarebout in that the issue of whether we can consider a pedagogical agent useful and capable of enhancing learning is too broad, since it depends on a variety of conditions and on specific pedagogical features that agents should have.

As regards affective learning, affective embodied agents (AEA) are specific artifacts designed with the ability of emotional expression with the aim of acting as affective pedagogical tutors in order to help students overcome negative affect, such as boredom or frustration, during a learning process (Kim, Baylor, & Shen, 2007). In another study, Guo and Goh (2016) incorporated an AEA into an information literacy game, finding that it can enhance students' motivation, enjoyment, perceived usefulness and behavioral intention. However, their AEA used minimal artificial intelligence; consequently, its affective expressions and feedback were not always sufficiently believable for the players-students. Other specific studies examined the effect of the emotional feedback created by an AEA on behavioral intention to use computer-based assessment (Terzis, Moridis, & Economides, 2012; van der Kleij, Eggen, Timmers, & Veldkamp, 2012). They concluded that different methods that provide feedback on students' learning outcomes may have different effects, which need further investigation. The design of AEAs is becoming increasingly sophisticated. However, research has focused more on how to deliver efficient information (a cognitive task) rather than on analyzing the metacognitive aspects of their use in educational systems.

Moreover, affective tutoring systems (ATS) combine affective computing techniques with emotional expressions in order to recognize learners' emotions during the learning process. Based on this information, they can provide appropriate emotional feedback in order to improve motivation, usability and interaction (Lin, Wang, Chao, & Chen, 2012; Wu, Huang, & Hwang, 2016). However, evidence about the effectiveness of affective pedagogical agents is still diverse and inconclusive. Most positive results have been based on experimental research in controlled learning environments. As a consequence, more research is needed in order to explore the behavior, efficiency and usefulness of affective pedagogical agents and affective feedback in authentic, long-term educational settings under different conditions, contexts, and learning situations (Arguedas, Daradoumis, & Xhafa, 2016b).

A considerable amount of research has investigated the role of pedagogical tutors and the metacognitive support they provide to students. In particular, in the field of self-regulated learning, Azevedo and Hadwin (2005) presented some initial challenges to the issue of scaffolding self-regulated learning and metacognition, which had specific implications for the design of computer-based scaffolds. Molenaar, van Boxtel, and Sleegers (2011) showed that using a pedagogical agent to support metacognitive activities resulted in improving students' metacognitive knowledge. Finally, it was seen that students who were provided with metacognitive support through a pedagogical agent developed better self-regulation skills (Karaoglan Yilmaz, Olpak, & Yilmaz, 2018). As a side effect of that study, the pedagogical agent's metacognitive support also had a significant effect on students' self-reflection skills. Boaler (2016) stressed the important role of self-reflection in making learners powerful by engaging them in a metacognitive process of thinking about what they know.

Finally, it is generally accepted that affective feedback can help students enhance self-regulation by informing them of what they did well (Labuhn, Zimmerman, & Hasselhorn, 2010). In this sense, affective feedback acts as metacognitive feedback, letting students know where they need to improve and what steps they can take to improve their work (Hattie & Timperley, 2007). This can help students not only improve their academic achievement (Brookhart, 2011) but also enhance their motivation (Wigfield, Klauda, & Cambria, 2011).

3. Research aims

3.1. Context

We performed a real experiment in a high school classroom setting based on a learning situation that involved specific metacognitive activities (which are described in more detail in the next section).

In this context, the Affective Pedagogical Tutor (APT) is a specifically designed agent that forms part of a larger framework comprising several components (Arguedas, Casillas, Xhafa, Daradoumis, Peña, & Caballé, 2016; Arguedas, Xhafa, Casillas, Daradoumis, Peña, & Caballé, 2018). This framework involves an emotion analysis model, which first analyzes text and conversation (wiki, chats and forum debates) generated by students involved in collaborative learning activities. It then proceeds to identify and represent the students' emotions that take place during these activities in a non-intrusive way.

This information is shown to both the human teacher and the APT, thus providing *emotion awareness* with regard to the way students' emotions emerge and evolve over time. This enables both the teacher and the APT to offer students affective feedback that influences students' motivation, engagement, self-regulation, and learning outcome.

In this study we explored students' self-reflection, referring to *what they learned* (what has changed with respect to their initial ideas and knowledge), *how they learned* (what led them to change their points of view), and which were the biggest difficulties they met. Accordingly, we set the following research questions.

3.2. Research questions

RQ1. In comparison to a human teacher's affective feedback, to what extent has an APT's affective feedback managed to enhance students' self-reflection?
RQ2. Which types of affective feedback proved to be more appropriate and effective for this learning situation?

3.3. Definition of variables for the learning situation

Both independent and dependent variables involved in the study are presented in Table 1. The learning situation contains metacognitive activities aimed at engaging students in a reflective learning process.

Table 1. Dependent and independent variables of the study

	Metacognitive activity
Independent variable:	A = Affective feedback
Dependent variable:	R = Student' self-reflection

4. Method

4.1. Materials

We designed a scenario which involved an authentic learning experience with high school students. The scenario included a main collaborative learning activity, "Design of a website," provided by the human teacher to instruct students in how to design a website following specific design principles. Designing a well-structured and consistent website is not a simple task, especially for inexperienced students, as in our case. Once the main learning activity was concluded, the teacher engaged students in a specific metacognitive learning situation, which is described below.

4.2. Learning situation: Metacognitive activity

Reflection is a learning process in itself that actively engages students to review the tasks they have carried out, to think about how they have performed in them, and ultimately how and what they have learned (Boud, 2001; Dewey, 1933). Indeed, when students reflect, they try to "focus on the cognitive aspects (thinking, problem solving, and so on) that led to particular actions, the outcomes and lessons learned from those actions, and how these inform what they might do in the future" (Mair, 2012, p. 148). Yet, the reflective process is "a complex one in which both feelings and cognition are closely interrelated and interactive" (Boud, Keough, & Walker, 1985, p. 11). There is a variety of research devoted to the study of the interactions and interrelations between metacognition and affective processes; this is evident especially in the area of self-regulated learning (e.g., Efklides, Schwartz, & Brown, 2018; Hudlicka, 2005). However, there are hardly any studies that examine the effect of emotional feedback provided by both the human teacher and a pedagogical agent on students' self-reflection during a metacognitive learning activity.

In this study, the learning situation comprised several online metacognitive activities that were carried out in the computer laboratory after the main class learning activities. Students were divided into small teams and the teacher created a chat space for each team in the Moodle platform and used specific questions and suggestions to encourage students to reflect upon the main learning activities they had performed in the class. The purpose of the online discussions in which students were engaged was threefold: to make students meditate on what they had learned (i.e., what had changed with respect to their initial beliefs), to understand how students learned (i.e., what led them to change their initial points of view), and to reflect on the difficulties they encountered during the realization of their collaborative activity and whether they dealt with them and how.

4.3. Participants and procedure

Participants were a sample of 45 fourth-year high school students attending the "Web Design" course. Within the sample, 11 of the students were girls (24%) and 34 were boys (76%). We randomly divided students into two big groups, a control and an experimental group, with 22 and 23 students respectively. In the control group, four teams were formed: two teams of five members and two teams of six members. In the experimental group, four teams were also formed: two teams of seven members, one team of five members and one team of four members. The teams were formed by the students themselves. Given that the synchronous online discussion lasted a maximum of one hour, we measured the student's emotional state after each student intervention in the chat.

The types of affective feedback provided are described in Table 2(a). They represent generic types of feedback, based on the theoretical model of feedback of Hattie and Timperley (2007). Since both the human tutor and the APT act independently, each provides their own particular feedback in their own wording and expression, that is, feedback articulation differs between the control and experimental groups. However, each particular feedback utterance should adhere to the generic feedback type it refers to. For the sake of illustration, we show some examples of affective feedback provided by either the teacher or the APT in the metacognitive activity.

The human teacher provides the following affective feedback type 8.1: “Do you remember how uncertain you were when you had to choose between different photos to represent the objectives of your page?”, whereas the APT provides the following affective feedback type 8.4: “Are you really happy you chose a single-page site approach instead of a blog-like homepage? Don’t you think that the latter could have provided your page with much more information?”

Table 2(b) presents the students’ conceptual and personal change, while Table 2(c) presents the students’ emotional states we considered to answer RQ2, based on Pekrun’s learning emotions (Pekrun et al., 2002). To answer RQ1, we considered the PAD (Pleasure-Arousal-Dominance) emotional state model (Mehrabian & O’Reilly, 1980).

Table 2(a). Affective feedback types that correspond to the metacognitive activity

Affective feedback types that support students’ self-reflection	
Make students reflect on the critical factors that influenced the realization of their learning activity	8.1
Make students think whether the type of feedback received during the learning activity was really helpful	8.2
Make students think about the information that would have been most appropriate to support their conceptual and personal change better	8.3
Make students meditate on alternative aspects that could have led them to take different decisions	8.4

Table 2 (b). Students’ conceptual and personal change

Students’ conceptual change (what students learned, i.e., what has changed with respect to their initial beliefs/knowledge)	
Make students think more critically about what they have learned in this course	6.1
Enable students to meditate that certain changes (in their knowledge and skills) evidently occurred with respect to what they initially thought or knew	6.2
Make students remember when these changes occurred	6.3
Make students think about what these changes are due to	6.4
Allow students to consider the aspects they are still confused about	6.5
Make students reflect on what they want to know more about	6.6
Students’ personal growth and understanding (how students learned, i.e., what led them to change their initial beliefs)	
Make students reflect on the actions they took to change their initial points of view	7.1
Let students remember what difficulties they have encountered that made it harder for them to achieve the desired changes	7.2
Enable students to meditate on how their perception was finally altered	7.3
Enable students to think about how their comprehension changed	7.4
Let students imagine how they are going to tackle their next work more efficiently	7.5

Table 2 (c). Students’ emotional states

E.1	Motivated
E.2	Curious
E.3	Confident
E.4	Pleased
E.5	Optimistic / challenging (stimulated)
E.6	Insecure or Embarrassed
E.7	Bored
E.8	Anxious or Dismayed
E.9	Outraged

It is worth mentioning here that we distinguish between two different ways of inferring emotion. The first is provided by our emotion awareness mechanism, which is used to identify the emotions that students experience during their work in the learning activities and which are retrieved through text (in our case, chat) analysis, as mentioned in the Context section above. The second way of inferring emotion is through the questionnaire, which contains questions related to specific emotional states (Table 2(c)) that students may experience when they receive affective feedback, either from the human teacher (control group) or the APT (experimental group). By responding to these specific questions, students basically evaluate the emotional effect that affective feedback types had on their self-reflection. That is, students’ self-reporting of affective states refers only to those affective states resulting from the affective feedback offered by the human teacher or the APT.

4.4. Data collection

This experiment supplied us with rich quantitative data, which enabled us to measure the effectiveness of the affective feedback types that the human teacher and the APT used in their group, as well as to evaluate the learners' emotional state with regard to the metacognitive learning situation.

The questionnaire was composed of:

- questions related to the affective feedback types presented in Table 2(a), that is, feedback types that support students' critical thinking;
- questions related to the students' individual conceptual change as well as students' personal growth and accountability, presented in Table 2(b);
- questions related to the different emotional states of students resulting from the affective feedback offered by the human teacher or the APT, shown in Table 2(c).

For all questions, we used a five-point Likert-type scale ranging from 1 (Almost never) to 5 (Almost always) and requiring a quantitative answer.

4.5. Reliability statistics and multivariate normality

Due to space restrictions, we provide a compact version of reliability statistics and multivariate normality measures rather than presenting them for each subscale. To ensure the reliability of data collection, Cronbach's alpha has been applied to both the control group (CG) and the experimental group (EG). The values of Cronbach's alpha obtained are shown in Table 3 and are higher than .70, thereby reinforcing the reliability of our indicators.

Table 3. Cronbach's alpha for the metacognitive learning activity in CG and EG

CG (N = 22)		EG (N = 23)	
Cronbach's alpha	No. elements	Cronbach's alpha	No. elements
.957	15	.915	15

In addition, the skewness and kurtosis of each variable were examined to check for multivariate normality. The critical values of all test statistics were calculated. The results showed that data were normally distributed as absolute values of skewness and kurtosis did not exceed the allowed maximum (2.0 for univariate skewness and 7.0 for univariate kurtosis).

5. Results

In this section, we present our results for the first research question through descriptive statistic measures.

5.1. The descriptive statistic measures

In this study we provide the most in-depth statistical analysis possible, in a gradual, progressive, and cumulative manner. Accordingly, the purpose of this section is to directly answer our first research question. The answer to the second research question is provided in detail in the Discussion section. To this end, we use descriptive statistic measures for comparing the two groups (control vs. experimental), thus evaluating the effectiveness of the APT's affective feedback with respect to that offered by the human teacher. As a consequence, we focus the analysis on a comparison of the two group's scores to check if there are any statistically significant differences in the effects of the different affective feedback types between the two groups.

5.2. The results with regard to metacognitive activity

With respect to the metacognitive learning activity, the questionnaire was composed of three parts, as seen in Table 4. With regard to *the items of the first part (6.1 – 6.6)*: The mean exceeded the value of three (3.0) in all items in both CG and EG. This indicates that all students (in both CG and EG) managed to carry out a fruitful meditation about what they had learned in this course and find out what has changed with respect to their initial ideas and knowledge. Certainly, EG students achieved better scores in this process. However, it is worth noting here the lower value of item 6.5 that EG students scored with respect to the CG students. This item refers to the process that students followed in order to think about and consider those aspects of the topic that they were still confused about. Personal interviews with EG students showed that the lower value that they obtained in item 6.5 was due to the fact that they did not consider such aspects since they did not need to. That is, the APT's affective feedback had managed to clarify things for them during the main learning activity.

With regard to *the items of the second part (7.1 –7.5)*: In the CG, the mean exceeded the value of three (3.0) in items 7.1-7.2 and 7.5, obtaining the values 3.53, 3.26, and 3.26 respectively. In EG, all values exceeded the value of three (3.0) in all items. This means that EG students managed to provide clear evidence of personal growth and understanding; that is, they were able to reflect on how they learned and what led them to change their initial points of view. In contrast, CG students demonstrated difficulties in meditating on how their perception and comprehension had eventually changed (items 7.3 and 7.4).

Table 4. The descriptive statistics in metacognitive learning activity

	(CG) ($N = 22$)				(EG) ($N = 23$)			
	Min	Max	Mean	*SD*	Min	Max	Mean	*SD*
6.1	1	6	3.16	1.608	1	6	4.27	1.077
6.2	1	6	3.79	1.475	1	6	4.27	1.241
6.3	1	6	3.11	1.629	2	6	4.00	1.345
6.4	1	5	3.05	1.545	2	6	4.00	1.380
6.5	1	6	3.79	1.653	1	6	3.32	1.644
6.6	1	6	3.37	1.674	2	6	4.32	1.171
7.1	1	6	3.53	1.541	2	6	4.09	1.306
7.2	1	5	3.26	1.727	1	6	3.95	1.253
7.3	1	5	3.00	1.700	2	5	3.82	1.220
7.4	1	5	2.89	1.595	1	6	4.27	1.077
7.5	1	6	3.26	1.790	2	6	4.59	1.436
8.1	1	5	2.95	1.747	1	5	4.00	1.414
8.2	1	6	3.58	1.575	1	6	3.95	1.463
8.3	1	6	3.11	1.912	1	6	4.00	1.690
8.4	1	6	3.47	1.744	1	6	4.36	1.529

Note. In all the tables, gray values indicate the best score obtained when we compare equivalent values in CG and EG.

With regard to *the items of the third part (8.1 –8.4)*: In the CG, the mean exceeded the value of three (3.0) in items 8.2-8.4, obtaining the values 3.58, 3.11 and 3.47 respectively. In EG, all values exceeded the value of three (3.0) in all items. Here again, EG students achieved high values of critical thinking skills in group work. CG students also showed quite acceptable similar skills with an exception in item 8.1 (they did not reflect so much on the critical factors that influenced the realization of the main learning activity).

Finally, in the last table, Table 5, it can be observed that all students (in both CG and EG) showed very similar feelings of pleasure, arousal, and dominance after the end of the learning scenario, with EG students being slightly more expressive about their personal satisfaction.

Table 5. Students' emotional states: The values obtained for pleasure, arousal and dominance in CG and EG

	CG ($N = 22$)				EG ($N = 23$)			
	Min	Max	Mean	SD	Min	Max	Mean	SD
Pleasure	5	6	5.49	.129	5.6	5.7	5.64	.0405
Arousal	5	5	4.99	.040	5.1	5.1	5.10	.0092
Dominance	5	5	5.05	.051	5.0	5.1	5.07	.0252

Note. In all the tables, gray values indicate the best score obtained when we compare equivalent values in CG and EG.

6. Discussion

In the previous section, the presentation of questionnaire results based on descriptive statistics provided us with insights about the effects of affective feedback types provided by both the human teacher and our virtual Affective Pedagogical Tutor (APT).

The purpose of this discussion is to focus on the experimental group and explore the types of affective feedback —used by our virtual Affective Pedagogical Tutor (APT) — which were more effective for improving students' self-reflection. As a side effect of this, we also draw some initial conclusions about the affective competencies the APT needs to have in order to achieve a positive change in students' conceptual, personal and affective aspects.

It is worth mentioning here that besides the importance of the feedback type itself, each feedback is a combination of gestural signals (emotional expressions), voice and/or text. Yet, the effectiveness of a feedback type is due mainly to the verbal power of the feedback rather than the non-verbal features of it.

Previous research on pedagogical agents yielded no difference (Heidig & Clarebout, 2011) or a small positive effect on cultivating student learning accountability (Schroeder, Adesope, & Gilbert, 2013). More recent research coincides in that the issue of whether we can consider a pedagogical agent useful and capable of enhancing learning is too broad, since it depends on a variety of conditions and on the specific pedagogical features that agents should have (Schroeder, Romine, & Craig, 2017). This also depends on the specific type of learning situation in which pedagogical agents try to be influential on learners' motivation or learning development (Dinçer & Doganay, 2017).

Taking previous research on pedagogical agents into account, we proceed to discuss and provide a response to the second research question of our metacognitive learning activity. We also take the opportunity to revisit and look at the first research question from the APT's point of view. To that end, we calculated the Pearson correlations of the different variables we defined (Tables 6 and 7) in order to identify the strong positive or negative linear relationships that exist among these variables. For the sake of convenience, we repeat each question below.

Table 6. Pearson correlations between APT's affective feedback and students' conceptual and personal change in experimental group (EG, $N = 23$)

	6.1	6.2	6.3	6.4	6.5	6.6	7.1	7.2	7.3	7.4	7.5
8.1	.517*	.211	.100	.022	.054	.387	-.154	.281	.412	.397	.538*
8.2	.291	.103	.711**	.535*	-.378	.210	.440	.309	.249	.446	.199
8.3	.555*	.402	.674**	.638**	.324	.508*	.546*	.227	.530*	.587**	.430
8.4	.665**	.603**	.353	.361	-.060	.622**	.336	.823**	.319	.438	.474*

Note. *Correlation is significant at the 0.05 level (2-tailed); **correlation is significant at the 0.01 level (2-tailed).

Table 7. Pearson correlations between APT's affective feedback and students' emotional states in experimental group (EG, $N = 23$)

	E.1	E.2	E.3	E.4	E.5	E.6	E.7	E.8	E.9
8.1	.478*	.515*	.374	.629**	.134	-.254	-.068	-.140	-.057
8.2	474*	.322	.305	.479*	.503*	-.287	-.380	-.143	-.415
8.3	.224	.329	.420	.520*	.363	.116	-.294	.278	-.106
8.4	.320	.298	.536*	.202	.585**	-.026	-.531*	-.070	-.343

Note. *Correlation is significant at the 0.05 level (2-tailed); **correlation is significant at the 0.01 level (2-tailed).

6.1. RQ1 and RQ2 regarding the metacognitive learning activity

RQ1. To what extent has an APT's affective feedback managed to enhance students' self-reflection?
RQ2. Which types of affective feedback proved to be more appropriate and effective for this learning situation?

In the experimental group (EG), the APT's affective feedback seems to have a positive effect on students' behavior. Indeed, the results in Table 6 indicate several strong positive relationships between all feedback types and students' conceptual and personal change to a greater or lesser extent. This finding is consistent with research indicating that that using a pedagogical agent to support metacognitive activities results in developing better metacognitive knowledge and self-regulation skills (Dignath & Büttner, 2008; Karaoglan Yilmaz, Olpak, & Yilmaz, 2018; Molenaar, van Boxtel, & Sleegers, 2011).

Unique to this study was the provision of different affective feedback types that focus on improving students' self-reflection on what and how they learned as well as the differentiation between the effectiveness of the different affective feedback types used. To the best of our knowledge, the effects of different types of affective feedback of a pedagogical agent in a real learning situation that involves metacognitive activities have not yet been investigated in a systematic study.

In our study, the affective feedback types with a major effect were 8.3 and 8.4; we therefore focus our discussion on these two. Indeed, APT feedback types 8.3 and 8.4 gave students the chance to think critically about the appropriateness of the information they received during the main activities. They also enabled students to provide insights about alternative aspects that, if they were supplied to them, would have led them to take different decisions.

The combination of feedback types 8.3 and 8.4 managed to make students think more critically about almost all elements we considered as basic influential factors for their conceptual and personal change (items 6.1, 6.2, 6.3, 6.4, 6.6, 7.1, 7.2, 7.3, 7.4 and 7.5 in Table 6, and described in Table 2(b)): what students learned, that is, what had changed with respect to their initial beliefs/knowledge; how students learned, that is, what led them to change their initial beliefs.

This trend echoes findings from other studies indicating that using different types of reflection prompts students to feel more engaged in metacognition by exhibiting different categories of metacognitive knowledge, such as planning learning tasks, monitoring comprehension, or evaluating progress (Menekse, 2020; Menz & Xin, 2016). Furthermore, our results are broadly consistent with previous research that has suggested that pedagogical agents can strengthen learners' reflection on what they have done or engaged in (Daumiller & Dresel, 2018) as well as improve the reasoning and decision-making abilities of their users (Le & Wartschinski, 2018). The research also underlines that the type of verbal feedback they provide really matters (Lin, Atkinson, Christopherson, Joseph, & Harrison, 2013).

In our study, there was only one element that had a non-influential relationship (and this occurred with all four APT affective feedback types): item 6.5 (*allow students to consider the aspects they were still confused about*). This is certainly not an easy matter to assess based only on quantitative data. As mentioned in the Results section, personal communication with EG students showed that these students simply did not consider the need to deal with such aspects, since at the end of all the main learning activities, they had completed the course goals successfully. They were also very happy at the end, as we can see in Table 7 (E.4). Yet, prompting students to reflect on confusing concepts lets them engage in a process for identifying the confusing concepts, while stimulating self-monitoring activities, such as comprehension reviews and searches for related knowledge (Menekse, 2020).
Based on the above, the answer to RQ1 is positive.

Regarding RQ2, as seen above, the types of affective feedback which proved to be more appropriate and effective for this learning situation were feedback types 8.3 and 8.4. As regards the other two APT affective feedback types (8.1 and 8.2), Table 6 shows that, though they offered some help to the students' critical thinking process, they certainly need to be further elaborated and improved.

Finally, Table 7 shows that the APT's affective feedback increased students' positive emotional states (E1 to E.5) at the end of the activity. However, this feedback did not have any significant relationship with students' negative emotional states (E6 to E.9), except feedback 8.4, which contributed to reducing students' boredom (E.7). We therefore need to further explore the reasons for this occurrence. That is, we need to examine why the APT's affective feedback did not have any influence on students' feelings of insecurity, anxiety, and anger during the metacognitive activity.

All in all, we are conscious that this study is the beginning of a complex and challenging endeavor and that more work still needs to be done in order to improve the APT design and ensure a truly worthwhile learning experience for students.

7. Conclusion

The study presented in this paper constitutes a real online educational experience involving secondary level students, a study which was missing in the fields of affective pedagogical tutors (APT) and metacognition and learning. So far, many agent-based studies have been laboratory-based and the participants were often university students. Unique to this study was the opportunity to examine different types of affective feedback used by the APT and determine which proved to be more appropriate and effective for cultivating reflective learning. The successful types of affective feedback employed by the APT in this work give an insight into the affective competencies the virtual APT needs to have in order to achieve a conceptual change in students as well as personal growth and understanding. Certainly, more research is needed to establish a more consolidated APT design with well-grounded and influencing affective competencies that could identify and tackle problems in different ways.

7.1. Limitations of the study and directions for future research

First, our Affective Pedagogical Tutor (APT) should be capable of dealing with more profound reflective and metacognitive learning issues. This requires a more intelligent pedagogical, emotional and technological design, endowed with further artificial intelligence techniques for emotion recognition and dialogue facilities for generating smooth affective feedbacks. In addition, our study with the APT could be extended to make use of more profound reflective and metacognitive learning theories and metacognitive self-regulation scales (Ku & Ho, 2010; Schellings & Van Hout-Wolters, 2011; Tock & Moxley, 2017).

Second, the results of our experiments on APTs' effectiveness are drawn from the users' perceptions. This is done post-experimentally by means of questionnaires. However, further real-time user signals should be captured by other techniques, such as sensors, and analyzed. This information can be fed into the APT to make its behavior more adaptive. It can also be used to cross-check the questionnaires.

Third, since a metacognitive learning activity constitutes an important part of a complete educational scenario, it is very important from an emotional point of view to foster a relationship of trust between the APT and the students, establishing a relationship of complicity between them. The APT should also nurture students' sense of cohesion and belonging to the class. This is related with the important issue of the affective competencies that the APT needs to have in order to achieve students' conceptual and personal enhancement. To that end, further research should focus on analyzing the most effective types of actions a human teacher carries out, adapting each task to the individual progress of each student while sustaining and managing their emotional states to favor their particular learning. Our aim is to endow our APT with these human affective competencies.

Finally, during the metacognitive activities, the APT should also be able to comment on the results obtained and contribute to a reflection and improvement process. The purpose of this is to make students both meditate on how new knowledge has been acquired and analyze which new cognitive and emotional skills were revealed and used to manage their emotions, and thus enhance their holistic development.

References

Arguedas, M., Daradoumis, T., & Xhafa, F. (2016a). Analyzing how emotion awareness influences students' motivation, engagement, self-regulation and learning outcome. *Educational Technology & Society, 19*(2), 87-103.

Arguedas, M., Daradoumis, T., & Xhafa, F. (2016b). Analyzing the effects of emotion management on time and self-management in computer-based learning. *Computers in Human Behavior, 63*, 517-529.

Arguedas, M., Casillas, L., Xhafa, F., Daradoumis, T., Peña, A., & Caballé, S. (2016). A Fuzzy-based approach for classifying students' emotional states in online collaborative work. In *Proceedings of the 10th International Conference on Complex, Intelligent and Software Intensive Systems* (pp. 61-68). Fukuoka, Japan: IEEE Computer Society.

Arguedas, M., Xhafa, F., Casillas, L., Daradoumis, T., Peña, A., & Caballé, S. (2018). A Model for providing emotion awareness and feedback using fuzzy logic in online learning. *Soft Computing, 22*, 963–977.

Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition – Implications for the design of computer-based scaffolds. *Instructional Science, 33*(5), 367–379.

Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology, 94*(2), 416-427.

Bahreini, K., Nadolski, R., & Westera, W. (2012). FILTWAM - A Framework for online affective computing in serious games. *Procedia CS, 15*, 45-52.

Belland B., Kim C. M., & Hannafin M. (2013). A Framework for designing scaffolds that improve motivation and cognition, *Educational Psychologist, 48*(4), 243–270.

Boaler, J. (2016). *Mathematical mindsets: Unleashing students' potential through creative math, inspiring messages, and innovative teaching.* San Francisco, CA: Jossey-Bass & Pfeiffer Imprints.

Boud, D., Keogh, R., & Walker, D. (1985). What is reflection in learning? In D. Boud, R. Keogh, & D. Walker (Eds.), *Reflection: Turning experience into learning* (pp. 7-17). New York, NY: Nichols Publishing.

Boud, D. (2001). Using journal writing to enhance reflective practice. *New Directions for Adult and Continuing Education, 2001*(90), 9-18.

Brookhart, S. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice, 30*(1), 3–12.

Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). Student engagement and student learning: Testing the linkages. *Research in Higher Education, 47*(1), 1–32.

Chinnery, S. A., Appleton, C., & Marlowe, J. M. (2019). Cultivating students' reflective capacity through group-based mindfulness instruction. *Social Work with Groups, 42*(4), 291-307.

Daumiller, M., & Dresel, M. (2018). Supporting self-regulated learning with digital media using motivational regulation and metacognitive prompts. *The Journal of Experimental Education, 42*, 1–16.

Desautel, D. (2009). Becoming a thinking thinker: Metacognition, self-reflection, and classroom practice. *Teachers College Record, 111*(8), 1997-2020.

Dewey, J. (1933). *How we think: A Restatement of the relation of refl ective thinking to the educative process.* Boston, MA: Heath.

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A Meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning, 3*(3), 231–264.

Dinçer, S., & Doganay, A. (2017). The Effects of multiple-pedagogical agents on learners' academic success, motivation, and cognitive load. *Computers & Education, 111*, 74-100.

Efklides, A., Schwartz, B. L., & Brown, V. (2018). Motivation and affect in self-regulated learning: Does metacognition play a role? In D. H. Schunk & J. A. Greene (Eds.), *Educational psychology handbook series. Handbook of self-regulation of learning and performance* (pp. 64–82). Washington, DC: Routledge/Taylor & Francis Group.

Elliott, C., Rickel, J., & Lester, J. (1999). Lifelike pedagogical agents and affective computing: An Exploratory synthesis. In M. Wooldridge, & M. Veloso (Eds.), *Artificial Intelligence Today, volume 1600 of Lecture Notes in Computer Science* (pp. 195-212). Berlin, Germany: Springer-Verlag.

Feidakis, M., Caballé, S., Daradoumis, T., Gañán, D. & Conesa, J. (2014). Providing emotion awareness and affective feedback to virtualized collaborative learning scenarios. *International Journal of Continuing Engineering Education and Life-Long Learning, 24*(2), 141-167.

Feidakis, M., Daradoumis, T., Caballé, S., Conesa, J., & Gañán, D., (2013). A Dual-modal system that evaluates user's emotions in virtual learning environments and responds affectively. *Journal of Universal Computer Science, 19*(11), 1638-1660.

Guo, Y. R., & Goh, D. H-L. (2016). Evaluation of affective embodied agents in an information literacy game. *Computers & Education, 103*(2016), 59-75.

Hattie, J., & Timperley, H. (2007). The Power of feedback. *Review of Educational Research, 77*, 81–112.

Heidig, S., & Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? *Educational Research Review, 6*, 27–54.

Henderson, K., Napan, K., & Monteiro, S. (2004). Encouraging reflective learning: An Online challenge. In R. Atkinson, C. McBeath, D. Jonas-Dwyer, & R. Philips (Eds.), *Beyond the comfort zone: Proceedings of the 21st ASCILITE Conference* (pp. 356–364). Perth, WA: ASCILITE.

Hudlicka, E. (2005). Modeling interactions between metacognition and emotion in a cognitive architecture. *Metacognition in Computation*, 55–61.

Jackson, C., & Larkin, M. J. (2002). Rubric: Teaching students to use grading rubrics. *Teaching Exceptional Children, 35*, 40–45.

Jarvela, S., Kirschner, P. A., Panadero, E., Malmberg, J., Phielix, C., Jaspers, J., Koivuniemi, M., & Jarvenoja, H. (2015). Enhancing socially shared regulation in collaborative learning groups: Designing for CSCL regulation tools. *Educational Technology Research and Development, 63*, 125–142.

Jennings, P. A. (2011). Promoting teachers' social and emotional competencies to support performance and reduce burnout. In A. Cohan & A. Honigsfeld (Eds.), *Breaking the mold of preservice and inservice teacher education: Innovative and successful practices for the twenty-first century* (pp. 133–143). New York, NY: Rowman & Littlefield.

Karaoglan Yilmaz, F. G., Olpak, Y. Z., & Yilmaz, R. (2018). The Effect of the metacognitive support via pedagogical agent on self-regulation skills. *Journal of Educational Computing Research, 56*(2), 159–180.

Kim, Y., Baylor, A.L. & Shen, E. (2007). Pedagogical agents as learning companions: The Impact of agent emotion and gender. *Journal of Computer Assisted Learning, 23*(3), 220-234.

Ku, K. Y. L., & Ho, I. T. (2010). Metacognitive strategies that enhance critical thinking. *Metacognition and Learning, 5*, 251–267.

Labuhn, A. S., Zimmerman, B. J., & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: the influence of feedback and self-evaluative standards. *Metacognition and Learning, 5*(2), 173–194.

Le, N. T., & Wartschinski, L. (2018). A Cognitive assistant for improving human reasoning skills. *International Journal of Human-Computer Studies, 117*, 45–54.

Lin, H.-C. K., Wang, C.-H., Chao, C.-J., & Chen, M.-K. (2012). Employing textual and facial emotion recognition to design an affective tutoring system. *Turkish Online Journal of Educational Technology, 11*, 418–426.

Lin, L., Atkinson, R. K., Christopherson, R. M., Joseph, S. S., & Harrison, C. J. (2013). Animated agents and learning: Does the type of verbal feedback they provide matter? *Computers & Education, 67*, 239–249.

Lin, X. (2001). Designing metacognitive activities. *Educational Technology Research and Development, 49*(2), 23-40.

Mair, C. (2012). Using technology for enhancing reflective writing, metacognition and learning. *Journal of Further and Higher Education, 36*(2), 147-167.

Mehrabian, A., & O'Reilly, E. (1980). Analysis of personality measures in terms of basic dimensions of temperament. *Journal of Personality and Social Psychology, 38*(3), 492-503.

Menekse, M. (2020). The Reflection-informed learning and instruction to improve students' academic success in undergraduate classrooms. *The Journal of Experimental Education, 88*(2), 183-199.

Menz, P., & Xin, C. (2016). Making students' metacognitive knowledge visible through reflective writing in a mathematics-for-teachers course. *Collected Essays on Learning and Teaching - Simon Fraser University, IX*, 155-166.

Molenaar, I., van Boxtel, C., & Sleegers, P. (2011). Metacognitive scaffolding in an innovative learning arrangement. *Instructional Science, 39*(6), 785–803.

Parkes, K. A., & Kajder, S. (2010). Eliciting and assessing reflective practice: A Case study in Web 2.0 technologies. International *Journal of Teaching and Learning in Higher Education, 22*(2), 218-228.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A Program of quantitative and qualitative research. *Educational Psychologist, 37*, 91–106.

Schellings, G., & Van Hout-Wolters, B. (2011). Measuring strategy use with self-report instruments: Theoretical and empirical considerations. *Metacognition and Learning, 6*, 83–90.

Schmitt, M. C. (1990). A Questionnaire to measure children's awareness of strategic reading process. *The Reading Teacher, 43*, 454–456.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334-370). New York, NY: Macmillan.

Schroeder, N. L., Adesope, O. O., & Gilbert, R. B. (2013). How effective are pedagogical agents for learning? A Meta-analytic review. *Journal of Educational Computing Research, 49*(1), 1-39.

Schroeder, N. L., Romine, W. L., & Craig, S. D. (2017). Measuring pedagogical agent persona and the influence of agent persona on learning. *Computers & Education, 109*, 176-186.

Shen, L., Wang, M., & Shen, R. (2009). Affective e-learning: Using "emotional" data to improve learning in pervasive learning environment. *Educational Technology & Society, 12*(2), 176–189.

Silver, N. (2013). Reflective pedagogies and the metacognitive turn in college teaching. In M. Kaplan, N. Silver, D. LaVaque-Manty, & D. Meizlish (Eds.), *Using reflection and metacognition to improve student learning* (pp. 1–17). Sterling, VA: Stylus Publishing.

Stone, B. A., & Lester, J. C. (1996). Dynamically sequencing an animated pedagogical agent. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence* (pp. 424–431). Portland, OR: The MIT Press.

Terzis, V., Moridis, C., & Economides, A. (2012). The Effect of emotional feedback on behavioral intention to use computer based assessment. *Computers & Education, 59*, 710–721.

Tock, J. L., & Moxley, J. H. (2017). A Comprehensive reanalysis of the metacognitive self-regulation scale from the MSLQ. *Metacognition and Learning, 12*, 79–111.

van der Kleij, F., Eggen, T., Timmers C., & Veldkamp, B. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education, 58*, 263–272.

Wigfield, A., Klauda, S. L., & Cambria, J. (2011). Influences on the development of academic self-regulatory processes. In B. J. Zimmerman, & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 33-48). New York, NY: Routledge.

Wu, C. H., Huang, Y. M., & Hwang, J. P. (2016). Review of affective computing in education/ learning: Trends and challenges. *British Journal of Educational Technology, 47*(6), 1304-1323.

Facilitating Students' Critical Thinking and Decision Making Performances: A Flipped Classroom for Neonatal Health Care Training

Ching-Yi Chang[1], Chien-Huei Kao[1] and Gwo-Jen Hwang[2*]
[1]Department of Midwifery and Women Health Care, National Taipei University of Nursing and Health Sciences, Taiwan // [2]Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan // frinng.cyc@gmail.com // chenhuei@ntunhs.edu.tw // gjhwang.academic@gmail.com
[*]Corresponding author

(Submitted April 5, 2020; Revised May 15, 2020; Accepted June 16, 2020)

ABSTRACT: In various case-based training courses, such as engineering, science and medical courses, students need to learn not only the skills to deal with problems, but also the knowledge to identify problems and make correct decisions. Such educational objectives have been recognized by educators as being important but challenging. In this study, an RSI (Recognize, Summarize, Inquire)-based flipped classroom is proposed to achieve this aim. Moreover, an explorative study was designed to probe the effectiveness of the proposed approach in a nursing case-based training course. The case-handling performances of the students trained with the RSI-based flipped classroom and those trained with the conventional flipped classroom were compared. The findings reveal that the RSI-based flipped classroom promoted the students' learning achievement, self-efficacy, critical thinking, and satisfaction more than the conventional case-based training did.

Keywords: Flipped classroom, Flipped learning, RSI, Critical thinking, Decision making

1. Introduction

In training courses aiming to foster students' competences of handling cases, such as engineering, science and medical courses, students need to learn not only the skills to deal with problems, but also the knowledge to identify problems and make correct decisions. A classical application of such courses is the training of nursing skills. For example, in the program of newborn health care, neonatal assessment is an important skill to ensure that nursing staff can provide proper treatments during the delivery of newborns (Fuloria & Kreiter, 2002; Heymann, 1993). In physical examination assessment, neonatal assessment can reduce potential disease of neonatal and enable early detection of neurological or developmental problems. It is used as the standard for first-line medical staff to assess whether a newborn needs follow-up intensive care or follow-up assessment (Sullivan, Miller, Fontaine, & Lester, 2012). It also involves evaluating Apgar scores, basic vital signs, the nervous system and fetal developmental maturity. Each step needs to be monitored to determine whether there are abnormal signs (Fuloria & Kreiter, 2002; Georgieff, 1995). When assessing the musculoskeletal and nervous system functions of newborns, medical staff must have sufficient expertise and skills to judge and deal with the changes in newborns' physical signs (Alexander & Kuo, 1997; Brodish, 1981). Therefore, improving medical staff's assessment ability is particularly important to help maintain the safety of newborns (Rüdiger & Aguar, 2012).

In the traditional neonatal assessment training course, simulation aids are used to give students opportunities to practice assessment skills (Kola & Bijapur, 2019; Solà-Pola et al., 2020; Yigzaw et al., 2019). Researchers have indicated that the difficulties and challenges in traditional neonatal assessment are mainly due to the insufficient time for teachers to explain the signs of rare diseases as well as to guide students to practice neonatal assessments (Blake, 2012; Tappero & Honeyfield, 2018). Although some scholars have applied flipped classrooms in nursing education by shifting the lectures to the pre-class time to enable teachers to have more time to guide students to practice in the class, it still remains a challenge to foster students' decision-making competence (Chang, Chang, Hwang, & Kuo, 2019). Scholars believe that one of the problems is the lack of effective learning guidance strategies to facilitate students' deep thinking in learning with the instructional videos in the before-class learning stage (Chang, Kao, Hwang, & Lin, 2019; Kirch, 2016). Abeysekera and Dawson (2015) further pointed out that the degree of students' involvement in pre-class learning has a great impact on their performance in the class. Thus, this study proposes a flipped classroom learning approach based on RSI.

To verify the usefulness of the RSI approach, this study implemented this model on an e-learning platform and conducted an experiment in a nursing school neonatal health care training course.

2. Literature review

2.1. Flipped classrooms

The flipped classroom is a blended learning mode in which students generally learn with instructional materials in an individual space before the class, so that teachers can guide them to practice, discuss, or apply knowledge in the group space (Bergmann & Sams, 2012; Bhagat, Chang, & Chang, 2016; Lo, Lie, & Hew, 2018). Scholars have reported that the flipped classroom enables teachers to schedule more time to engage students in applying knowledge, which can promote their high-order thinking and communication competences owing to more interactions with teachers and peers for sharing knowledge and ideas (Chang, Chang, Hwang, & Kuo, 2019; Hwang & Lai, 2017). In recent years, the popularization of mobile systems with wireless networks has facilitated students' reading of the instructional materials in the pre-class stage, as well as assisting teachers in conducting in-class activities to promote interactions and knowledge sharing among students (Hsia & Sung, 2020). In addition, scholars have incorporated a large number of learning strategies or tools into flipped classrooms to improve students' learning achievements (Awidi & Paynter, 2019; Chang, Chang, Hwang, & Kuo, 2019). Several scholars have confirmed the effectiveness of flipped classrooms from various perspectives, such as allowing students to learn in an interactive and autonomous manner (Liou, Bhagat, & Chang, 2016; McLaughlin & Rhoney, 2015; Mirriahi, Alonzo, McIntyre, Kligyte, & Fox, 2015; Miles, Lee, Foggett, & Nair, 2017), improving students' learning performances (Dehghanzadeh & Jafaraghaee, 2018; Hwang, Lai, & Wang, 2015; Peterson, 2016; Wang, 2017), and promoting their self-efficacy (Tawfik & Lilly, 2015).

In the past decade, flipped classrooms have gradually been adopted in medical and nursing education with promising outcomes (Critz & Knight, 2013), such as improving students' learning achievements and learning smartification in emergency training courses for newborns (Rose, Claudius, Tabatabai, Kearl, Behar, & Jhun, 2016), promoting a sense of responsibility and problem-solving skills in a psychiatric nursing practicum course (Lee, Chang, & Jang, 2017), and improving critical thinking in a musculoskeletal medical-surgical nursing theoretical training course (Dehghanzadeh & Jafaraghaee, 2018). In addition, researchers have stated the potential of flipped classrooms in promoting learners' higher order thinking (Alsowat, 2016; Hussein et al., 2019). For example, Lee (2018) reported that students had better critical thinking performance in flipped scientific reading activities than those in traditional classrooms; Ding, Li, and Chen (2019) reported that flipped classrooms could promote students' critical thinking by employing a proper learning design in an ophthalmology course. Lin (2019) conducted an experiment in a software engineering course and indicated that the provision of proper supports in flipped classrooms could facilitate their problem-solving performance.

Despite a number of successful examples, scholars have pointed out several issues to be addressed when implementing flipped classrooms. For example, students tend to watch instructional videos without deep thinking in the pre-class stage, which is likely to lead to shallow discussion or poor performance in class (Lei, Yau, Lui, Tam, Yuen, & Lam, 2019; Luo, Kushnazarov, & Hew, 2019; Parsons & Beauchamp, 2012). Basal (2015) pointed out that, to implement a successful flipped classroom, it is necessary to adopt appropriate strategies to facilitate students' deep thinking when learning with the instructional videos in the pre-class stage. This is particularly crucial in case-based training courses, which not only aim to have learners memorize and comprehend the learning content, but also fosters their competences of making decisions and solving problems (Danielson & Berntsson, 2007; Phillips, 2000). Therefore, incorporating effective strategies into the individual space of flipped classrooms to improve students' learning performance and higher order thinking is a crucial issue (Kirch, 2016; Shannon, 2008; Yilmaz & Baydas, 2017). Several scholars have also reported similar concerns in flipped classrooms (Lin & Hsia, 2019; Lin, Hwang, & Hsu, 2019; Zhang, Fan, Xia, Guo, Jiang, & Yan, 2017).

2.2. Strategies and pedagogical theories of flipped classrooms

The literature has shown the importance of the pre-class stage in flipped classrooms. Several researchers have indicated that students' pre-class learning status could affect their in-class learning outcomes, including the learning tasks related to problem solving and critical thinking (Huong, Huy, & Ha, 2018; Li, 2019). Kirch (2012) further emphasized the need to guide students to learn in the pre-class stage, including guiding them to take notes, summarizing the learning content and raising questions. Several flipped classroom studies have also focused on improving students' pre-class performances in different courses, such as nursing skills training (Lin, Hwang, & Hsu, 2019; Mudd & Silbert-Flagg, 2016; Zhu, Lian, & Engström, 2020).

On the other hand, scholars have stated the need to further guide students to think in depth and to make reflections in addition to taking notes, summarizing learning content and raising questions (Lin & Hsia, 2019). From the perspective of identifying and solving problems, such as case handling in nursing education, it is important to guide students to have in-depth thinking and make reflections, so that they can correctly make clinical judgments as well as mastering clinical skills (Asselin, Schwartz-Barcott, & Osterman, 2013; Hicks-Moore & Pastirik, 2006). By referring to the experiential learning theory by Kolb (1976), it is important to facilitate students' learning through concrete experience, reflective observation, abstract conceptualization and active experimentation. In nursing training courses, those case studies generally provide a form of concrete experience to learners, while taking notes, answering questions and completing learning sheets could be treated as a form of abstract conceptualization. This indicates that additional guidance is needed to help learners reflect and explore based on what they have learned.

Thus, the study emphasized the significance of guiding students to think in-depth in a step-by-step manner when learning with instructional videos in the pre-class stage through recognizing key concepts (i.e., "Recognize"), organizing what they have learned (i.e., "Summary"), thinking in depth to inquire about the potential problems (i.e., "Inquire"). To verify the usefulness of the RSI-based flipped classroom approach, an experiment was conducted in a neonatal health care training course to answer the following research questions:

- Can the RSI-based flipped approach better improve the nursing students' learning achievement in neonatal assessment than the conventional flipped approach?
- Can the RSI-based flipped approach better improve the nursing students' self-efficacy in neonatal assessment than the conventional flipped approach?
- Can the RSI-based flipped approach better enhance the nursing students' critical thinking in neonatal assessment than the conventional flipped approach?
- Can the RSI-based flipped approach better enhance the nursing students' problem solving ability in neonatal assessment than the conventional flipped approach?
- Can the RSI-based flipped approach better enhance the nursing students' learning satisfaction in neonatal assessment than the conventional flipped approach?

3. Developing the RSI-based flipped learning environment

3.1. System structure

The RSI-based flipped learning environment was implemented on the Tronclass platform. Figure 1 shows the system structure. The teacher interface enables teachers to maintain student profiles or learning portfolios, edit test items, and design instructional videos and learning sheets. Students can use smartphones or tablets to view the learning materials and learning sheets, take notes, and complete learning tasks.

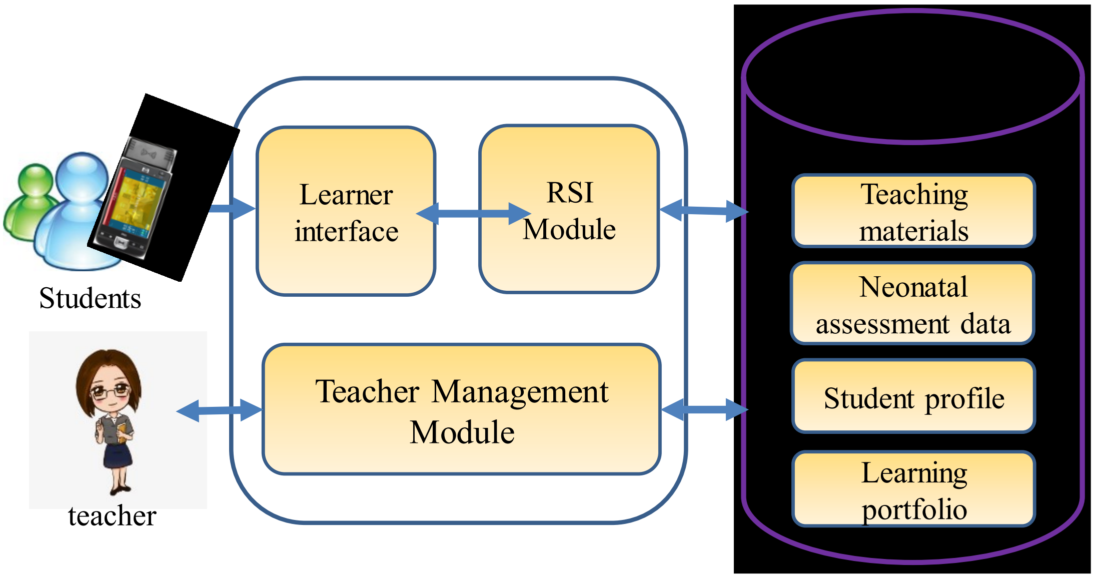


Figure 1. System structure

Figure 2. Interface for browsing the neonatal assessment instructional videos

*Figure 3***.** Interface for annotating and summarizing the learning content

Students can use smartphones to access the learning system to watch the instructional videos and complete the learning tasks. The RSI-based learning mode consists of three stages. In the first stage, students recognize the key problems and take notes by watching instructional videos. Figure 2 shows the learner interface for a nursing training program: the instructional videos of neonatal assessment.

In the second stage, students are guided to summarize what they have learned and identified, as presented in Figure 3.

In the third stage, students are guided to inquire based on what they have summarized and identified. In this stage, they are encouraged to make reflections on the learning process, think diversely, and find potential problems regarding what they have learned. Figure 4 shows the interface of guiding students to reflect, think, and raise questions.

Figure 4. Interface for guiding students to reflect, think and question

4. Method

The experiment was conducted at a teaching hospital in Taiwan for a neonatal assessment training program.

4.1. Subjects

The subjects were 36 nursing students from two classes of a nursing university in northern Taiwan. Their average age was 21. One class with 18 students was chosen as the experimental group learning with the RSI-based flipped classroom. Another class with 18 students was the control group learning with the conventional flipped approach. The two classes received the same learning content and were instructed by the same teacher.

4.2. Experimental design

The schedule of the experiment is shown in Figure 5. Before the activity, the two groups were administered a pre-test and pre-questionnaires. During the experiment, the two groups learned through the online instructional videos in the pre-class time. In this stage, the experimental group learned with the RSI-based flipped approach via the online learning system to complete the learning task. That is, in the pre-class stage, the students were guided by the RSI procedure to recognize the key concepts in the instructional videos, summarize what they had learned, propose questions, and inquire by in-depth thinking or seeking evidence. On the other hand, the control group learned with the conventional flipped classroom. In the conventional mode, the students were asked to watch videos and take notes, do some exercises, and complete a learning sheet prepared by the teacher by answering a set of questions based on what they had learned from the learning content. After the activity, the students were administered a post-test and post-questionnaires.

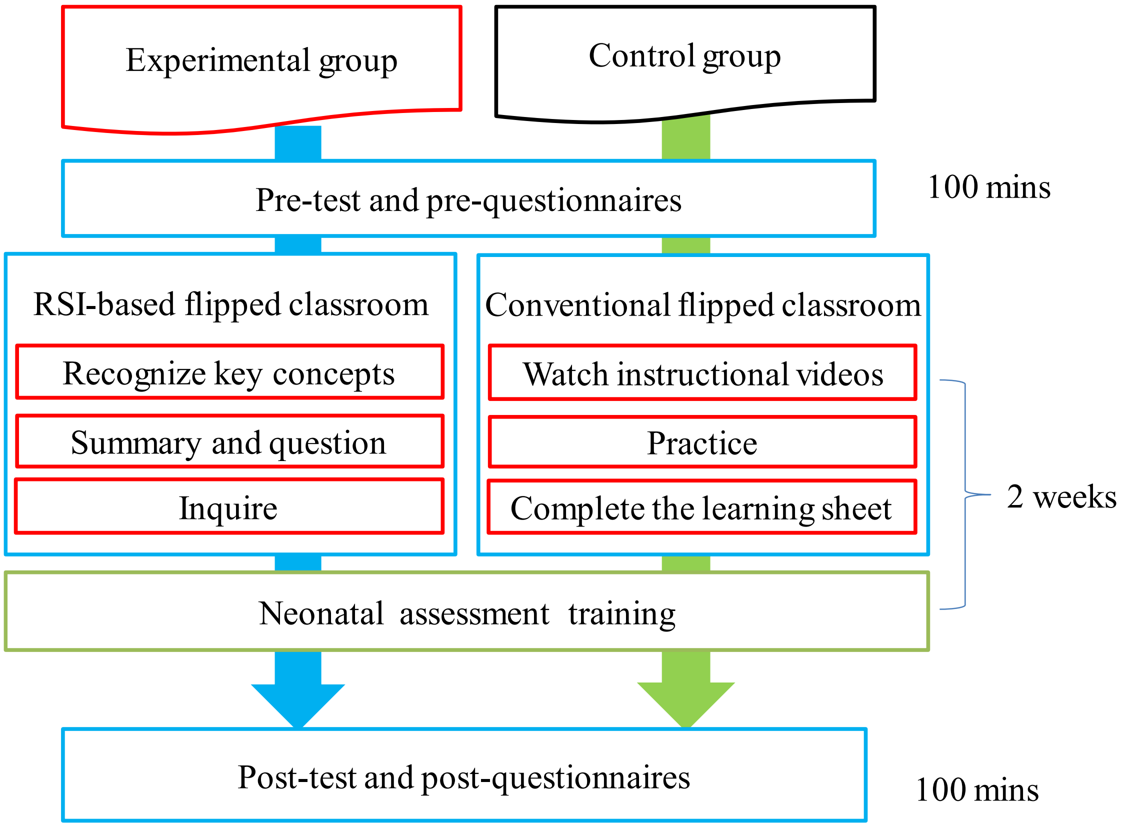


Figure 5. Procedure of the experimental design

In the class, the students were guided by the teacher to make decisions on several medical cases as well as practice relevant nursing skills. An example is shown in Figure 6, in which the students were asked to deal with a newborn case and to verbally explain the assessment information, such as swallowing reflex, gag reflex, rooting reflex, sucking reflex, Moro reflex, startle reflex, tonic neck reflex, Babinski's reflex, grasping reflex, stepping reflex, and crawling reflex.

Figure 6. In-class learning activity: students practice the nursing skills and make decisions during the OSCE process

4.3. Measuring tools

The pre-test was designed by two nursing teachers who had taught the course for more than 12 years. It was composed of 20 multiple-choice questions for evaluating learners' knowledge of clinical neonatal assessment. The perfect score was 100.

Table 1. Rubrics for the neonatal assessment skills test

Dimension	Items to observe
Vital Signs	heart rate, PR, BT
Distress	Facial expression, respiratory effort, activity, tone
Color	Tongue, mucous membranes (centrally pink vs. cyanotic), nail beds
Nutrition status	Subcutaneous fat, breast nodule
Hydration status	Skin turgor, anterior fontanel
Gestational age	Skin (smooth vs. peeling), ear cartilage, areola and nipple formation, breast nodule, sole creases, descent of tests, rugage, labia
Neurologic status	Posture, tone, activity, response to stimuli, cry, state, nipples (number and position), skin color
Respiratory/chest status	Respiratory rate and effort, retractions, nasal flaring, grunting, audible stridor or wheezing, chest shape, nipples (number and position), skin color
Cardiovascular status	Precordial activity, visible point of maximal intensity, skin perfusion and color
Abdomen	Size (full, distended, taut, shiny), shape (round, concave), distention (generalized or localized), visible peristaltic waves, visible bowel loops, muscular development/tone
Head	Size, shape, fontanels, suture lines, swelling, hair distribution, condition of hair
Eyes	shape, size, position, pupils, blink, extraocular movements, color of sclera, discharge, ability to fix and follow
Ears	shape, position, external auditory canal, response to sound
Nose	shape, nares, flaring, nasal bridge
Mouth	shape, symmetry, movement, philtrum, tongue
Neck	shape, range of motion, webbing, masses
Genitalia (male)	Scrotum, descent of testes, rugae, inguinal canals, foreskin, penile size, urine stream, meatus, perineum, anus
Genitalia (female)	Labia majora, labia minora, clitoris, vagina, perineum, inguinal canals, anus
skin	Color, texture, firmness, vernix caseosa, masses, lanugo, lesions (pigmentary, vascular, trauma-related, infectious)
Extremities	Posture, range of motion

The Rubrics for neonatal assessment skills tests originate from the physical assessment of neonatal postnatal performance proposed by Tappero and Honeyfield (2018). It consists of 20 dimensions, as shown in Table 1. Each dimension is scored with a 3-point rating scheme for evaluating learners' OSCE skills: 3 means completely meeting the standard, 2 means partially meeting the standard, and 1 means that the operation is incorrectly performed. Two experienced teachers were recruited to evaluate students' case-handling performances using the rubrics.

The self-learning efficacy scale was modified based on the scale developed by Pintrich et al. (1991). It consists of eight items, such as "I am confident that I can understand the most complex parts taught by teacher" and "I am confident that I can learn the key concepts taught by teachers." A 5-point Likert scoring scale was adopted in this measure, and the Cronbach's α value was .93.

The learning satisfaction scale was proposed by Chu et al. (2010). It is composed of nine items, such as "Using this way to learn, I can make some new discoveries or new knowledge" and "Using this way can help me learn to distinguish things." A 5-point Likert scoring scale was adopted in the measure. Its Cronbach's α value was .91.

The problem-solving questionnaire was proposed by Hwang and Chen (2017). A total of five items are included, such as "When solving a problem, I try to identify the problem type first" and "Before solving a problem, I think I need to understand the cause of the problem." A 5-point Likert scoring scale was adopted and its Cronbach's α value was .78.

The critical thinking scale was proposed by Hwang and Chen (2017). It consists of five items, such as "I ask myself periodically if I am meeting my goals" and "I periodically review to help me understand important relationships." A 5-point Likert scoring scale was adopted and its Cronbach's α value was .83.

4.4. Data analysis

To analyze the neonatal assessment OSCE scores as well as their self-efficacy, critical thinking, problem-solving, and learning satisfaction of the students learning with different approaches (i.e., the RSI-based flipped classroom and conventional flipped classroom), ANCOVA (Analysis of Covariance) was adopted. The Shapiro-Wilk test results of the ratings for individual measures are between 0.88 and 0.90 ($p > .05$). This reveals that data for individual scales have a normal distribution.

5. Results

5.1. Neonatal assessment OSCE result

To evaluate the students' neonatal assessment skills, ANCOVA was utilized to compare the neonatal assessment skills test scores of the two groups. The neonatal assessment skills test was the dependent variable and the pre-test was the covariate. The Levene's test of variance showed that the homogeneity assumption was confirmed with $F(1, 34) = .66$ ($p > .05$). In addition, the homogeneity of regression slopes was $F(1, 32) = .59$ ($p > .05$). Therefore, the ANCOVA could be conducted.

The ANCOVA result is shown in Table 2. The adjusted means and *SD* values were 96.67 and 1.59 for the experimental group, and 83.81 and 1.59 for the control group. The post-test scores of the two groups were significantly different with $F(1, 33) = 25.31$ ($p < .001$). The experimental group had significantly higher scores than the control group. Furthermore, the effect size (η^2) of learning approach was .434, which indicated a 1arge to medium effect size.

Table 2. ANCOVA results on students' neonatal assessment OSCE

Groups	*N*	Mean	*SD*	Adjusted mean	Adjusted *SD*	*F*	η^2
Experimental group	18	95.63	7.67	96.67	1.59	25.31^{***}	.434
Control group	18	82.78	5.75	83.81	1.59		

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

5.2. Self-efficacy

ANCOVA was utilized to compare the self-efficacy survey of the two groups. The self-efficacy post-test was the dependent variable and the pre-test was the covariate. The Levene's test showed that the homogeneity assumption was confirmed with $F(1, 34) = .38$ ($p > .05$). In addition, the homogeneity of regression slopes was $F(1, 32) = 1.46$ ($p > .05$). Therefore, the ANCOVA could be conducted.

The ANCOVA result is shown in Table 3. The adjusted means and *SD* values were 3.83 and 0.11 for the experimental group, and 3.28 and 0.11 for the control group. The post-test scores of the two groups were significantly different with $F(1, 33) = 5.05$ ($p < .05$). The experimental group had significantly higher scores than the control group. Furthermore, the effect size (η^2) of learning approach was .133, which indicated a medium effect size.

Table 3. ANCOVA results on students' self-efficacy

Groups	*N*	Mean	*SD*	Adjusted mean	Adjusted *SD*	*F*	η^2
Experimental group	18	3.68	.62	3.83	.11	5.05^{*}	.133
Control group	18	3.13	.34	3.28	.11		

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

5.3. Critical thinking

To evaluate the students' critical thinking, ANCOVA was used to analyze the critical thinking survey of the two groups. The critical thinking post-test was the dependent variable and the pre-test was the covariate. The

Levene's test revealed that the homogeneity assumption was confirmed with $F(1, 34) = .47$ ($p > .05$). In addition, the homogeneity of regression slopes was $F(1, 32) = 3.74$ ($p > .05$). Therefore, ANCOVA could be conducted.

The analysis result is shown in Table 4. The adjusted means and *SD* values were 4.55 and 0.13 for the experimental group, and 2.74 and 0.13 for the control group. The post-test scores of the two groups were significantly different with $F(1, 33) = 62.45$ ($p < .001$). The experimental group had significantly higher scores than the control group. Furthermore, the effect size (η^2) was .753, revealing a large to medium effect size.

Table 4. ANCOVA Results on students' critical thinking

Groups	*N*	Mean	*SD*	Adjusted mean	Adjusted *SD*	*F*	η^2
Experimental group	18	4.53	.50	4.55	.13	62.45^{***}	.753
Control group	18	2.75	.60	2.74	.13		

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

5.4. Problem-solving

ANCOVA was used to analyze the problem-solving survey of the two groups. The problem-solving post-test was the dependent variable and the pre-test was the covariate. The Levene's test showed that the homogeneity assumption was confirmed with $F(1, 34) = .17$ ($p > .05$). In addition, the homogeneity of regression slopes was $F(1, 32) = 1.57$ ($p > .05$). Therefore, ANCOVA could be applied.

The ANCOVA result is shown in Table 5. The adjusted means and *SD* values were 4.49 and 0.13 for the experimental group, and 2.77 and 0.13 for the control group. The post-test scores of the two groups were significantly different with $F(1, 33) = 86.80$ ($p < .001$). The experimental group had significantly higher scores than the control group. Furthermore, the effect size (η^2) was .725, revealing a large to medium effect size.

Table 5. ANCOVA results on students' problem-solving ability

Groups	*N*	Mean	*SD*	Adjusted mean	Adjusted *SD*	*F*	η^2
Experimental group	18	4.47	.50	4.49	.13	86.80^{***}	.725
Control group	18	2.77	.60	2.77	.13		

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

5.5. Learning satisfaction

To evaluate the students' learning satisfaction, ANCOVA was employed. The learning satisfaction post-test was the dependent variable and the pre-test was the covariate. The Levene's test showed that the homogeneity assumption was confirmed with $F(1, 34) = .16$ ($p > .05$). In addition, the homogeneity of regression slopes was $F(1, 32) = 2.99$ ($p > .05$). Therefore, ANCOVA could be applied.

The ANCOVA result is shown in Table 6. The adjusted means and *SD* values were 4.60 and 0.14 for the experimental group, and 2.80 and 0.14 for the control group. The post-test scores of the two groups were significantly different with $F(1, 33) = 87.25$ ($p < .001$). The experimental group had significantly higher scores than the control group. Furthermore, the effect size (η^2) was .726, revealing a 1arge to medium effect size.

Table 6. ANCOVA results on students' learning satisfaction

Groups	*N*	Mean	*SD*	Adjusted mean	Adjusted *SD*	*F*	η^2
Experimental group	18	4.59	.49	4.60	.14	87.25^{***}	.726
Control group	18	2.81	.67	2.80	.14		

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

6. Conclusions and implications

In this study, an RSI-based flipped approach was proposed and implemented in a case-handling nursing course, Neonatal Assessment. The findings in the experiment reveal that the proposed approach has great potential for improving students' neonatal assessment performance, self-efficacy, critical thinking, problem-solving ability and learning satisfaction.

6.1. Research question 1: Can the RSI-based flipped approach better improve the nursing students' learning achievement in neonatal assessment than the conventional flipped approach?

In terms of learning performance, it is inferred that the RSI-based approach facilitated the subjects' in-depth thinking in the pre-class time of the flipped classroom, which can further affect their performance in the corresponding in-class stage. As indicated by several flipped learning studies, acquiring knowledge in an effective and in-depth manner is the foundation for better involving students in the follow-up activities in flipped classrooms (Helgevold & Moen, 2015; Post, Deal, & Hermanns, 2015).

From the perspective of constructivism, engaging students in raising rather than answering questions raised by the teachers after browsing and summarizing the learning content enables them to reexamine the content from different aspects and to attempt to connect the new knowledge with their prior knowledge or past experience (Chien, Chen, & Liao, 2019; Jong, Chen, Tam, & Chai, 2019). This reveals that RSI not only facilitated the students' in-depth thinking, but also their attempt to extend their learning scope and to reorganize what they had learned, and therefore it helped the students make correct decisions to complete their learning goals by gaining and organizing knowledge in an effective manner. The findings are in line with those reported by several past studies regarding the use of question-proposing strategies, such as Lin and Hsia (2019) and Lin, Hwang, and Hsu (2019).

6.2. Research question 2: Can the RSI-based flipped approach better improve the nursing students' self-efficacy in neonatal assessment than the conventional flipped approach?

The experimental results also show that the RSA-based approach promoted the students' self-efficacy, which refers to a person's belief that he/she can successfully complete certain tasks or achieve certain desired goals (Bandura, 1988). Using the RSI mechanism, the students were guided not only to watch the instructional videos, but also to organize and reexamine what they had learned. It is inferred that students' self-efficacy was promoted owing to the fact that they had the opportunity to know the whole picture regarding the learning content and to think in depth by exploring relevant information. This echoes the findings of Hsia and Hwang (2020), that is, effective teaching strategies stimulate students' potential and improve their self-efficacy.

6.3. Research question 3: Can the RSI-based flipped approach better enhance the nursing students' critical thinking in neonatal assessment than the conventional flipped approach?

In terms of high-order thinking, which has been indicated as a potential objective of flipped classrooms by several scholars (Chang, Chang, Hwang, & Kuo, 2019; Ha, O'Reilly, Ng, Zhang, & Serpa, 2019), it is deduced that the RSI-approach encouraged the students to think from diverse perspectives when they were guided to raise questions in the pre-class time. Several previous studies regarding question-proposing also reported that, when trying to raise questions, students generally search for more relevant information and try to view the learning content from different perspectives; moreover, they are more willing to discuss with peers regarding the learning topics to gain more opinions from different perspectives (He, Holton, & Farkas, 2018; Ziegelmeier & Topaz, 2015). This implies that the students not only viewed the learning content in diverse ways, but also had more opportunities to resolve cognitive conflicts, which could contribute to the result that their critical thinking was promoted.

6.4. Research question 4: Can the RSI-based flipped approach better enhance the nursing students' problem solving ability in neonatal assessment than the conventional flipped approach?

The experimental results show that the participants who learned with the mechanism of the RSI-based flipped approach outperformed those who learned via the conventional flipped approach in terms of problem solving. In the application of the present study, through the RSI procedure, students can evaluate new born babies' status, taking into account diverse perspectives, such as Vital Signs, Distress, Color, etc. This allows them to interpret and organize the information, raise questions and think in depth based on the whole picture they have. This finding is consistent with the research of Lin, Hwang, and Hsu (2019); that is, through the provision of stepwise guidance in flipped classrooms, students' problem solving ability could be improved.

6.5. Research question 5: Can the RSI-based flipped approach better enhance the nursing students' learning satisfaction in neonatal assessment than the conventional flipped approach?

The learning satisfaction results show that the participants who learned with the mechanism of the RSI-based flipped approach outperformed those who learned via the normal flipped approach. This indicates that the integration of the RSI-based flipped approach strategy into neonatal assessment activities can effectively improve students' learning satisfaction. In the meantime, from the post-test results, it was found that the experimental groups showed significantly better learning satisfaction than the control group, implying that the challenges of the learning tasks and the increased complexity of the learning materials were at an appropriate level within the zone of proximal development proposed by Vygotsky (1978). This finding complies with what has been reported by Lin and Hsia (2019) and Lin, Hwang, and Hsu (2019) that step-by-step guiding can inspire students to learn as it also increases their deep learning in the meantime.

6.6. Limitations and suggestions

It should be noted however that there are some limitations in the present study. First, its results were mainly derived from quantitative analysis; to further investigate the factors affecting students' learning performances and perceptions, it would be better to conduct in-depth interviews in the future. Second, owing to the low birth rate in Taiwan in the past decades, the sample size was not large, implying the need to perceive the findings in a conservative manner. Third, the application of the present study is neonatal assessment, and hence the research results can only be applied to those nursing or medical training programs with similar aims and features.

From the findings and discussion in this study, several suggestions for future research are given as follows:

(1) Examining the impacts of personal factors when using the RSI-based flipped approach. The factors could be students' learning performance, personal characteristics or perceptions. By taking the factors into consideration, more precise suggestions can be provided to help teachers and researchers use the RSI approach in better ways.
(2) Probing the effectiveness of the RSI-based flipped classroom from different angles. It is recommended that researchers should combine the convenience of science and technology and focus on analyzing more key factors that affect students' learning effects, such as applied technology learning resources, the content of course activities, multimedia formats, etc.
(3) Providing instant feedback to individual studies during the case-handling process in training programs. It is suggested that researchers can consider implementing a more customized and convenient online learning system using artificial intelligence or other new technologies to support RSI-based flipped classrooms in a more effective way.
(4) Probing the factors affecting students' learning outcomes through the RSI-based flipped classroom approach. It would be valuable to find the factors that stimulate students to think more deeply, improve their critical thinking ability, and promote their high-level reflection performance in the pre-class stage to promote their decision-making performances in dealing with the training cases in the class.
(5) Applying the RSI-based flipped approach to other training programs aiming at fostering students' decision-making and problem-solving competences. Engineering courses and scientific inquiries as well as other nursing or medical courses could be potential applications.

In summary, learning design plays a crucial role in education research and learner-centered learning. In case-handling training courses, such as nursing skills training, the RSI-based flipped classroom learning could be promising for those courses aiming at training students' competences of dealing with cases. As a consequence, it could be crucial to apply the approach to other courses and collect more data to further evaluate its usefulness in the future. It is also crucial to probe whether using the RSI approaches in other learning modes, such as inquiry-based mobile learning or contextual learning will have similar results.

Acknowledgement

This study was funded by the Ministry of Science and Technology of Taiwan. The project ID was MOST-108-2511-H-011-005-MY3. It was also funded by the National Taipei University of Nursing and Health Sciences under contract number 109ntunhs-NT-05 and the 109 Teacher Growth Community Program Subsidy.

References

Abeysekera, L., & Dawson, P. (2015). Motivation and cognitive load in the flipped classroom: Definition, rationale and a call for research. *Higher Education Research & Development*, *34*(1), 1-14.

Alexander, M., & Kuo, K. N. (1997). Musculoskeletal assessment of the newborn. *Orthopedic nursing*, *16*(1), 21-31.

Alsowat, H. (2016). An EFL flipped classroom teaching model: Effects on English language higher-order thinking skills, student engagement and satisfaction. *Journal of Education and Practice, 7*(9), 108-121.

Asselin, M. E., Schwartz-Barcott, D., & Osterman, P. A. (2013). Exploring reflection as a process embedded in experienced nurses' practice: A Qualitative study. *Journal of Advanced Nursing*, *69*(4), 905-914.

Awidi, I. T., & Paynter, M. (2019). The Impact of a flipped classroom approach on student learning experience. *Computers & Education*, *128*, 269-283.

Bandura, A. (1988). Organizational application of social cognitive theory. *Australian Journal of Management, 13*(2), 275–302.

Basal, A. (2015). The Implementation of a flipped classroom in foreign language teaching. *Turkish Online Journal of Distance Education*, *16*(4), 28-37.

Bergmann, J., & Sams, A. (2012). Before you flip, consider this. *Phi Delta Kappan*, *94*(2), 25-25.

Bhagat, K. K., Chang, C. N., & Chang, C. Y. (2016). The Impact of the flipped classroom on mathematics concept learning in high school. *Educational Technology & Society*, *19*(3), 134-142.

Blake, D. (2012). Newborn examination: The Student's role? *British Journal of Midwifery*, *20*(12), 892-896.

Brodish, M. S. (1981). Perinatal assessment. *Journal of Obstetric, Gynecologic, & Neonatal Nursing*, *10*(1), 42-46.

Chang, B. Y., Chang, C. Y., Hwang, G. H., & Kuo, F. R. (2019). A Situation-based flipped classroom to improving nursing staff performance in advanced cardiac life support training course. *Interactive Learning Environments*, *27*(8), 1062-1074.

Chang, C. Y., Kao, C. H., Hwang, G. J., & Lin, F. H. (2019). From experiencing to critical thinking: a contextual game-based learning approach to improving nursing students' performance in Electrocardiogram training. *Educational Technology Research and Development*, 1-21.

Chien, C. F., Chen, G. Y. H., & Liao, C. J. (2019). Designing a connectivist flipped classroom platform using unified modeling language. *International Journal of Online Pedagogy and Course Design*, *9*(1), 1-18.

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.

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

Critz, C. M., & Knight, D. (2013). Using the flipped classroom in graduate nursing education. *Nurse Educator*, *38*(5), 210-213.

Danielson, E., & Berntsson, L. (2007). Registered nurses' perceptions of educational preparation for professional work and development in their profession. *Nurse education today*, *27*(8), 900-908.

Dehghanzadeh, S., & Jafaraghaee, F. (2018). Comparing the effects of traditional lecture and flipped classroom on nursing students' critical thinking disposition: A Quasi-experimental study. *Nurse Education Today*, *71*, 151-156.

Ding, C., Li, S., & Chen, B. (2019). Effectiveness of flipped classroom combined with team-, case-, lecture-and evidence-based learning on ophthalmology teaching for eight-year program students. *BMC Medical Education*, *19*(1), 419-428.

Fuloria, M., & Kreiter, S. (2002). The Newborn examination: Part I. Emergencies and common abnormalities involving the skin, head, neck, chest, and respiratory and cardiovascular systems. *American Family Physician*, *65*(1), 61-68.

Georgieff, M. K. (1995). Assessment of large and small for gestational age newborn infants using growth curves. *Pediatric Annals*, *24*(11), 599-607.

Ha, A. S., O'Reilly, J., Ng, J. Y., Zhang, J. H., & Serpa, S. (2019). Evaluating the flipped classroom approach in Asian higher education: Perspectives from students and teachers. *Cogent Education*, *6*(1), 1638147.

He, W., Holton, A. J., & Farkas, G. (2018). Impact of partially flipped instruction on immediate and subsequent course performance in a large undergraduate chemistry course. *Computers & Education*, *125*, 120-131.

Helgevold, N., & Moen, V. (2015). The Use of flipped classrooms to stimulate students' participation in an academic course in initial teacher education. *Nordic Journal of Digital Literacy*, *10*(01), 29-42.

Heymann, M. (1993). Early postnatal assessment of the newborn in the developing country: Malawi. *The ABNF Journal*, *4*(4), 90-94.

Hicks-Moore, S. L., & Pastirik, P. J. (2006). Evaluating critical thinking in clinical concept maps: A Pilot study. *International Journal of Nursing Education Scholarship*, *3*(1). doi:10.2202/1548-923X.1314

Hsia, L. H., & Hwang, G. J. (2020). From reflective thinking to learning engagement awareness: A Reflective thinking promoting approach to improve students' dance performance, self-efficacy and task load in flipped learning. *British Journal of Educational Technology*. doi:10.1111/bjet.12911

Hsia, L. H., & Sung, H. Y. (2020). Effects of a mobile technology-supported peer assessment approach on students' learning motivation and perceptions in a college flipped dance class. *International Journal of Mobile Learning and Organisation*, *14*(1), 99-113.

Huong, H. T. L., Huy, N. H. D., & Ha, N. N. (2018). The Flipped classroom: Using thematic teaching to develop critical thinking for high school students. *American Journal of Educational Research*, *6*(6), 828-835.

Hussein, M. H., Ow, S. H., Cheong, L. S., & Thong, M. K. (2019). A Digital game-based learning method to improve students' critical thinking skills in elementary science. *IEEE Access*, *7*, 96309-96318.

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., & Lai, C. L. (2017). Facilitating and bridging out-of-class and in-class learning: An Interactive E-book-based flipped learning approach for math courses. *Educational Technology & Society, 20*(1), 184-197.

Hwang, G. J., Lai, C. L., & Wang, S. Y. (2015). Seamless flipped learning: A Mobile technology-enhanced flipped classroom with effective learning strategies. *Journal of Computers in Education*, *2*(4), 449-473.

Jong, M. S. Y., Chen, G., 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.

Kirch, C. (2016). *Flipping with Kirch: The ups and downs from inside my flipped classroom*. New Berlin: The Bretzmann Group.

Kola, D. A., & Bijapur, M. S. (2019). A Study to evaluate the effectiveness of simulation on newborn care among III year B. Sc. Nursing students of SDM Institute of Nursing Sciences, Sattur, Dharwad, Karnataka. *International Journal of Nursing Science Practice and Research*, *5*(1), 64-69.

Lee, Y. H. (2018). Scripting to enhance university students' critical thinking in flipped learning: Implications of the delayed effect on science reading literacy. *Interactive Learning Environments*, *26*(5), 569-582.

Lee, M. K., Chang, S. J., & Jang, S. J. (2017). Effects of the flipped classroom approach on the psychiatric nursing practicum course. *Journal of Korean Academy of Psychiatric and Mental Health Nursing*, *26*(2), 196-203.

Lei, C. U., Yau, C. W., Lui, K. S., Tam, V., Yuen, A. H., & Lam, E. Y. (2019). Designing instructional videos and classwork activities: Teaching internet of things via flipped classroom. *International Journal of Mobile Learning and Organisation*, *13*(4), 392-411.

Li, L. (2019). An Action research of O2O blended learning in the integrated English class under the context of a Chinese private language university. *The Online Journal of Distance Education and e-Learning*, *7*(4), 261-268.

Lin, H. C., Hwang, G. J., & Hsu, Y. D. (2019). Effects of ASQ-based flipped learning on nurse practitioner learners' nursing skills, learning achievement and learning perceptions. *Computers & Education*, *139*, 207-221.

Lin, Y. N., & Hsia, L. H. (2019). From social interactions to strategy and skills promotion. *Educational Technology & Society*, *22*(2), 71-85.

Lin, Y. T. (2019). Impacts of a flipped classroom with a smart learning diagnosis system on students' learning performance, perception, and problem solving ability in a software engineering course. *Computers in Human Behavior, 95*, 187-196.

Liou, W. K., Bhagat, K. K., & Chang, C. Y. (2016). Beyond the flipped classroom: A Highly interactive cloud-classroom (HIC) embedded into basic materials science courses. *Journal of Science Education and Technology*, *25*(3), 460-473.

Lo, C. K., Lie, C. W., & Hew, K. F. (2018). Applying “First Principles of Instruction” as a design theory of the flipped classroom: Findings from a collective study of four secondary school subjects. *Computers & Education*, *118*, 150-165.

Luo, J., Kushnazarov, M., & Hew, K. F. (2019). An Analysis of undergraduate level flipped courses based on the seven principles: A Case study. *International Journal of Mobile Learning and Organisation*, *13*(4), 412-451.

McLaughlin, J. E., & Rhoney, D. H. (2015). Comparison of an interactive e-learning preparatory tool and a conventional downloadable handout used within a flipped neurologic pharmacotherapy lecture. *Currents in Pharmacy Teaching and Learning*, *7*(1), 12-19.

Miles, C. A., Lee, A. C., Foggett, K. A., & Nair, B. K. (2017). Reinventing medical teaching and learning for the 21st century: Blended and flipped strategies. *Archives of Medicine and Health Sciences*, *5*(1), 97-102.

Mirriahi, N., Alonzo, D., McIntyre, S., Kligyte, G., & Fox, B. (2015). Blended learning innovations: Leadership and change in one Australian institution. *International Journal of Education and Development using ICT*, *11*(1), 4-16.

Mudd, S. S., & Silbert-Flagg, J. (2016). Implementing the flipped classroom to enhance nurse practitioner clinical decision-making in the care of the pediatric asthma patient. *Nursing Education Perspectives*, *37*(6), 352-353.

Parsons, J., & Beauchamp, L. (2012). *From knowledge to action: Shaping the future of curriculum development in Alberta*. Edmonton, AB, Canada: Alberta Education.

Peterson, D. J. (2016). The Flipped classroom improves student achievement and course satisfaction in a statistics course: A Quasi-experimental study. *Teaching of Psychology*, *43*(1), 10-15.

Phillips, J. R. (2000). Rogerian nursing science and research: A Healing process for nursing. *Nursing science quarterly*, *13*(3), 196-201.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor, MI: The University of Michigan.Post, J. L., Deal, B., & Hermanns, M. (2015). Implementation of a flipped classroom: Nursing students’ perspectives. *Journal of Nursing Education and Practice*, *5*(6), 25-30.

Rose, E., Claudius, I., Tabatabai, R., Kearl, L., Behar, S., & Jhun, P. (2016). The Flipped classroom in emergency medicine using online videos with interpolated questions. *The Journal of Emergency Medicine*, *51*(3), 284-291.

Rüdiger, M., & Aguar, M. (2012). Newborn assessment in the delivery room. *Neoreviews*, *13*(6), e336-e342.

Shannon, S. V. (2008). Using metacognitive strategies and learning styles to create self-directed learners. *Institute for Learning Styles Journal*, *1*(1), 14-28.

Solà-Pola, M., Morin-Fraile, V., Fabrellas-Padrés, N., Raurell-Torreda, M., Guanter-Peris, L., Guix-Comellas, E., & Pulpón-Segura, A. M. (2020). The Usefulness and acceptance of the OSCE in nursing schools. *Nurse Education in Practice*, *43*, 102736. doi:10.1016/j.nepr.2020.102736

Sullivan, M. C., Miller, R. J., Fontaine, L. A., & Lester, B. (2012). Refining neurobehavioral assessment of the high-risk infant using the NICU network neurobehavioral scale. *Journal of Obstetric, Gynecologic & Neonatal Nursing*, *41*(1), 17-23.

Tappero, E. P., & Honeyfield, M. E. (2018). *Physical assessment of the newborn: A Comprehensive approach to the art of physical examination*. New York, NY: Springer Publishing Company.

Tawfik, A. A., & Lilly, C. (2015). Using a flipped classroom approach to support problem-based learning. *Technology, Knowledge and Learning*, *20*(3), 299-315.

Vygotsky, L. (1978). *Mind in society: The Development of higher psychological process.* Cambridge, MA: Harvard University Press.

Wang, F. H. (2017). An Exploration of online behaviour engagement and achievement in flipped classroom supported by learning management system. *Computers & Education*, *114*, 79-91.

Yigzaw, M., Tebekaw, Y., Kim, Y. M., Kols, A., Ayalew, F., & Eyassu, G. (2019). Comparing the effectiveness of a blended learning approach with a conventional learning approach for basic emergency obstetric and newborn care training in Ethiopia. *Midwifery*, *78*, 42-49.

Yilmaz, R. M., & Baydas, O. (2017). An Examination of undergraduates’ metacognitive strategies in pre-class asynchronous activity in a flipped classroom. *Educational Technology Research and Development*, *65*(6), 1547-1567.

Zhang, C., Fan, H., Xia, J., Guo, H., Jiang, X., & Yan, Y. (2017). The Effects of reflective training on the disposition of critical thinking for nursing students in China: A Controlled trial. *Asian Nursing Research*, *11*(3), 194-200.

Zhu, L., Lian, Z., & Engström, M. (2020). Use of a flipped classroom in ophthalmology courses for nursing, dental and medical students: A Quasi-experimental study using a mixed-methods approach. *Nurse Education Today*, *85,* 104262. doi:10.1016/j.nedt.2019.104262

Ziegelmeier, L. B., & Topaz, C. M. (2015). Flipped calculus: A Study of student performance and perceptions. *Primus*, *25*(9-10), 847-860.

Aljojo, N. (2020). Understanding the Sequence of Learning in Arabic Text -- Saudi Arabian Dyslexics and Learning Aid Software. *Educational Technology & Society*, *23* (2), 47–60.

Understanding the Sequence of Learning in Arabic Text -- Saudi Arabian Dyslexics and Learning Aid Software

Nahla Aljojo
Faculty of Computing and Information Technology, Information Systems Department, King Abdulaziz University, Jeddah, Saudi Arabia // University of Jeddah, College of Computer Science and Engineering, Department of Information Systems and Technology, Jeddah, Saudi Arabia // naljojo@kau.edu.sa // nmaljojo@uj.edu.sa

(Submitted December 23, 2019; Revised March 5, 2020; Accepted June 7, 2020)

ABSTRACT: Dyslexic individuals have serious difficulties in learning to read, and several software programs have been developed to overcome them. Previous research studies found that in most cases, providing correct spelling and word recognition with greater accuracy constitutes the key function. However, more research has to be conducted on the software programs for dyslexia. Another issue is visual-spatial attention; the aspect of directing attention to a location of what should be learnt while reading has been ignored in most studies. This research analysis of the previous studies outlines some research gaps as follows: recording children's eye movements during reading were uncertain. Another research gap is understanding the frequencies of Arabic word recognition for dyslexic children the meaning of the Arabic word during reading for dyslexic were not adequately presented in the previous research. Therefore, this paper presents an experimental study on identifying Arabic words and their meaning by dyslexic individuals focusing on the sequence of interactions. A decision-making analysis of the key interactive components of the Arabic alphabet puzzle app for dyslexics is performed, and the finding of the study reveals that the speed and accuracy of visual word recognition and meaning description increases and significantly influences visual short-term memory The analysis of the most important factors that influence all the other factors of the user experience of the Arabic alphabet puzzle app for dyslexics shows that 'desirability' constitutes the key attribute in enhancing visual-spatial attention and accelerating word recognition and meaning description for Arabic-speaking dyslexics.

Keywords: Eye Tracking, Reading, Dyslexia, Visual word, Visual stimuli

1. Introduction

Dyslexia constitutes a specific learning disability. It is defined as an impairment in the acquisition of reading and spelling abilities that manifests in neurodevelopmental reading difficulties (Reid, Shaywitz, & Shaywitz, 2013). One of the most common symptoms of the disorder is related to the difficulties experienced in word recognition. Dyslexia is caused by various genetic and environmental risk factors as well as their interaction (Peterson, and Pennington, 2015), although some studies have revealed that a final understanding of the causes of dyslexia or agreement on its precise definition has yet to be reached (Benfatto, Seimyr, Ygge, Pansell, Rydberg, & Jacobson, 2016). Its prevalence is attested by the fact that certain individuals experience persistent problems in learning to read for no apparent reason. Since reading constitutes a primary area of learning, this is consistently problematic. It may be the case that one individual struggles with reading and spelling while another has difficulty in understanding verbal communication. Dyslexia issues manifest in various aspects of learning. The major research question of this study focuses on whether software programs offer a solution for dyslexia. A number of practical approaches have been developed to assist dyslexic individuals based on certain software programs. Among the common software applications available in the public domain are "4 Pics 1 Word – Free," "A+ Spelling Test," "A1 Spelling App," "ABC Magic Phonics," "Leo-Recorder and Transcriptions," and "Reading Machine Free." However, in terms of aiding learning difficulties in reading, several issues have remained unclear and certain specific learning processes have been ignored such as the fact that it constitutes an assembling the sequence of events. In respect to the complexity of co-ordination in learning, sequencing and organizational difficulties with individual with dyslexia may benefit from certain technologies that allow them to concentrate on the content rather than the process (Smythe, 2010).

Even though dyslexics behave normally and are obviously intelligent, their condition always manifests in their difficulty to acquire reading abilities. This is what is clearly known about dyslexia for which several remedies have been proposed. Among the most feasible remedies is the use of educational counselling and tutoring. Unfortunately, this approach involves certain drawbacks, especially in terms of language rules and writing system. Considering that dyslexia is more prominent in children (Adlof & Hogan, 2018), adults also are affected. Its signs can be recognized as taking a long time before talking, difficulty in learning new words faster, extreme difficulty in framing words correctly, confusing words that sound alike, and difficulty in remembering or naming letters, numbers and colours. These attributes can be corrected by software programs that may come in the form of games or any interactive software programs.

The majority of available software applications for aiding dyslexics in general mostly relies on the so-called "recency effect" meaning an order of learning effect that relies on reflection of recent presentation (Cockburn et al., 2015). However, the "serial position effect" may aid dyslexic individuals more as it allows them to remember what has been learnt first and the last thing that was learned in a series (Corina, 2015). This falls under the category of "visual-spatial attention." In this study it is described as a form of directing attention to a location of what should be learnt during reading. Visual-spatial attention has been identified as a crucial predictor of reading abilities (Gabrieli & Norton, 2012). It has been suggested as an early way of identifying dyslexia, especially in pre-schoolers and predicts future reading acquisition. Considering this in the context of Arabic language acquisition, the Arabic alphabetic puzzle game application using eye tracking and chatbot proposed in Aljojo et al. (2018) is used in this study in order to examine how eye tracking can be used to improve the reading ability of dyslexic individuals when learning how to read. Visual-centric issues in software programs for educating and tutoring dyslexia remain a major concern. Individual learning behaviour is based on the intention, which in turn is influenced by the learning attitude. Visual properties are crucial to learning any software program. Some studies revealed that reading and object recognition (colour vision and high acuity) takes about 1.5° area responsible for sharp central vision (the fovea) (Bodis-Wollner, 2013). Visual effect is very important as a minimum of 3–4 and a maximum of 10 to 12 letters during reading appears visible in detail at any time during reading for a normal individual (Rayner et al., 2011). Research gaps were established based on the fact that previous studies ignored subjective evaluation of dyslexics and focus more software base intervention. Furthermore, the previous studies also disregarded recording children's eye movements during reading by observations. Another research gap is understanding the frequencies of Arabic word recognition with a Software program. The question whether the high visual outlook of a software program is the solution to dyslexics can be studied and answered by creating an experimental setting. Although experiments are known to alter the participant's responses and even create bias, they are useful in extracting facts and produce a sufficiently high amount of data. The current study relies on the user's perception in order to obtain statistically valid quantitative results through Arabic word recognition, finding and fixing missing puzzle as well as identifying the meaning of the word. The outcome of the study can be generalized with strong justifications. Hence, the proposed research questions are: How is dyslexics children's eye movements influence their reading comprehension? How is dyslexics children's frequencies of Arabic word recognition influence their reading comprehension.

The remaining part of the paper is organized as follows: Section 2 presents the related work; Section 3 describes and explains the research methodology; Section 4 offers the experimental analysis; Section 5 summarizes the results; and Section 6 completes the study with the conclusion and recommendation.

2. Related work

The application that has been developed was based on Arabic vowels and the graphical objects with eye tracker and chatbot to monitor and motivate dyslexic individuals. This complete technology package offers an efficient platform to increase the learning abilities of dyslexics (Aljojo et al., 2018). It is aimed at children aged four to seven who have dyslexia, including symptoms such as excessive movement and dispersed attention. Individuals who are unable to read normally due to their disability can also use this type of application to facilitate their learning process. This application is a puzzle-based game that utilizes eye tracking and chatbot to ensure that the user does not lose focus and keeps attention. The implementation process relies on the users' perception of how the game will work and how the application will provide helpful information that will support and simplify their needs.

Previous related studies have discussed various techniques for developing software/hardware suitable for individuals with dyslexia. Video games have been used for dyslexia correction. It was revealed that software video games have gained considerable popularity as they provide an enlarged size field of view and an adequate sequence of visual display (Green & Bavelier, 2003). Interestingly, it was also found out that the beneficial effects of certain kinds of software video game displays helped increase reading abilities (Franceschini et al., 2015). Another study has indicated that children with dyslexia who used a software video game showed an improvement in visuo-spatial and temporal attentional shifting matched with an improvement in reading speed without any increase in reading errors rate (Franceschini et al., 2015).

Individuals with dyslexia are given the opportunity to detect errors while reading which might improve their skills attributable to the very high degree of spelling-sound consistency (Bavelier et al., 2013). Many dyslexia educational and tutoring aided software programs available in the public domain are intended to achieve precisely that. The A1 Spelling App focuses on sound by engaging the learner to listen to vocabulary being spoken and repeat after the software program. ABC Magic Phonics uses a combination of a photo for each phonetic presented alongside with suggestions. Leo-Recorder and Transcriptions software program record and transcribe. Reading Machine Free is software program that shows how a word sounds and also how letters are combined from taken from the user's voice. The A+ Spelling Test software program allows the user to listen to a word he or she then spells and corrects it if needed. Shah (2012) also proposed a learning app suitable for dyslexia (1000 Sight Words Superhero). It is a software program developed for comparative word sets that involve 1,000 common words and their known counterparts. These words are organized in such a way that they can be matched to each other and their synonyms. Learners start with a set of 10 words and match them with other words that have the same meaning. Even though the program allows for an organized way of learning vocabulary, the level of provided interaction and action towards learning correct spelling and word recognition with greater accuracy has not been taken into account in this app. Lotum (2017) has recently developed a software program that uses images within puzzles to aid individual learners to identify correct words. The users point out right words following a set of rules as they progress. Unfortunately, visual-spatial attention and acceleration of recognition has not been taken into account in this app.

While the majority of the previous studies focused on developing full-fledged dyslexia tutoring apps, the analysis of visual content accuracy has not always been considered. It has been found that in textual search interfaces the dyslexic user's performance increases when icons are included in the search user interfaces (Berget et al., 2016). Text presentation within dyslexia tutoring app has been identified as a crucial element. Specifically, font types without serifs are considered as the most preferable for dyslexics (Evett & Brown, 2005) that also affect their reading performance (Rello & Ballesteros, 2015). Another study has revealed that dyslexics generally prefer designs with large text and images (Williams & Hennig, 2015), yet it is still unclear whether these conditions influence performance. This is also supported with the findings of Al-Wabiletal (2007), which indicated that dyslexics faced some enormous navigational drawbacks online, even on a platform with large text and images. In the context of visual spatial attention, Gabrieli and Norton (2012) studied the importance of visual spatial attention and observed that it constituted one of the major factors that influence reading abilities. Typically, the left side of visual space affect s the visual-spatial attention and it is the cause of abnormal function in reading behaviors (Nagamatsu et al., 2009). Franceschini et al. (2012) revealed that visual-spatial attention predicts the reading capabilities and the future performance in reading acquisition in preschoolers. Other studies reveal various conditions of the effect of Eye movements (Clifton et al., 2016), eye movement on word length and frequency on fixation (Kuperman & Van Dyke, 2011), eye movement orientation (Chang et al., 2016), and eye-tracking (Paracha et al., 2018) and (Frutos-Pascual & Garcia-Zapirain, 2015) with respect to reading ability.

It is furthermore worth mentioning that the aforesaid works (Aljojo et al., 2018; Shah, 2012; Green & Bavelier, 2003; Franceschini et al., 2012; Franceschini et al., 2015; Bavelier et al., 2013) focused on interactive practices of learning reading. However, how to employ eye movement or eye tracking on learning reading for practical applications has not been greatly discussed. Hence, following the previous successful applications of eye movement and word skipping during reading (Rayner et al., 2011), and the success of visual-spatial attention in reading (Franceschini et al., 2012 & Gabrieli & Norton, 2012), this paper presents the findings on how visual-spatial attention influences visual accuracy for dyslexics.

3. Methodology

The research methodology adopted for this study involves visual-spatial attention and multi-decision criteria analysis. The visual attention experiment was concerned with finding the missing puzzle, any change-detection from the way the participant gazed, were observed. The behavior of the participant's eye-movement on change-detection accuracy of finding the right puzzled was measured. The experiment established how the spatial locations occupied visual space of missing puzzle when they were no longer visible. The visual-spatial attention experiment was preceded by the multi-decision criteria analysis (Jasri & Rahim. 2017; Carpitella et al., 2018), decision-making trial and evaluation laboratory (DEMATEL) (Chang et al., 2011; Lee et al., 2013; Zhou et al., 2017). This was carried out in order to examine the key interaction component of visual-spatial attention of an eye tracking Arabic alphabetic puzzle app for dyslexia. While the visual-spatial attention experiment was set out to reveal the visual accuracy for dyslexics reading skills, the DEMATEL focused on investigating the key attributes of visual-spatial attention of the app. These methods were chosen based on the discrepancies of research on dyslexia tutoring software programs' nature of adoptions.

3.1. Visual-spatial attention experiment

This paper constitutes an extension of the work of Aljojo et al. (2018) who proposed an eye tracking and chatbot Arabic alphabetic puzzle game application for dyslexia. The application consists of three levels, namely a group of similar Arabic letters, vowels, and a play exercise. The play exercise page contains a gaming and learning activity. If the participant answers correctly, he/ she can access a game that is related to the same letter. The idea for developing the application was due to the lack of the sequence of interaction provided for Arabic word recognition, finding and fixing a scrambled puzzle, and identifying the meaning the words generated. Hence, a review was conducted for some available applications in the public domain.

3.1.1. Participants

A sample of 42 Saudi Arabian dyslexic children was used for the control experiment of this study. The 42 children used for the control experiment of this study are all in control group and, their parents are involved in the experimental, as a result only one group experiment was analyzed in order to avoid epistemological bias since parents are involved. 14 of them were males and the remaining 28 were females. The participant mean age was 9.12 years (range 8.1–13.9). All the children were recruited through their parents from different regions of Saudi Arabia. They are all Arabic natives and their literacy level in Arabic is at Primary school level the participants' parents completed a full review of the experimental details and were given the software program from Aljojo et al. (2018). Following the parents' consent to have their children participate in this study, two important criteria were set out as follows: The children had not been exposed to the software program from Aljojo et al. (2018) and were not aware of it; until the parents acquired the knowledge of the apps and then performed the prior experiment to lead them have knowledge about the content provided by the app before using the it. All participants were confirmed as diagnosed with dyslexia with no history of neurological disease.

3.1.2. App

The Arabic Alphabetic Puzzle Game Using Eye Tracking and Chatbot for Learning Disability (Dyslexia) (Aljojo et al., 2018) was used for the experiment (see Figure 1). Figure 2 presents the text base content that directly displays the Arabic letters with puzzles, three Arabic letter words supposedly constituting the minimum number of letters any normal individual can capture while reading (Rayner et al., 2011).

3.1.3. Research instrument and data collections

The pre-field experimental study was carried out with the participants two to four days before running the main experiment and follow-up experiment between the first and the last day, similar to some previous studies (Franceschini et al., 2013; Gori et al., 2013). The software program of Aljojo et al. (2018) developed for a mobile device was run for the participant. The matching puzzles were set and the missing words are completed. Each child was individually treated for three sessions for three days over a period of three weeks.

In terms of tasks administration and evaluation, all reading performances of the participants were recorded and time and errors were coded by Native-Arabic speakers. This works by observations, during the reading experiment (See Figure 1). It is an ethnographical observational exercises, specifically, covert observational approach in which the participants are not aware of the recording of their reading performance. The performance is defined as accurately reading correctly while the frequency of eye movement recorded. The errors are defined as the inaccuracies of reading, involving mismatch, mispronunciation of words, skipping of words. The researchers are Native-Arabic, they are aware of the accurately recording of the participants' reading performance, based on Rayner et al. (2013) who revealed that the potential utility of recording children's eye movements during reading when conducting future applied research is not by technology, it is by care repeated overt observation.

Figure 1. Group of Arabic letters to be identified and read by the participants

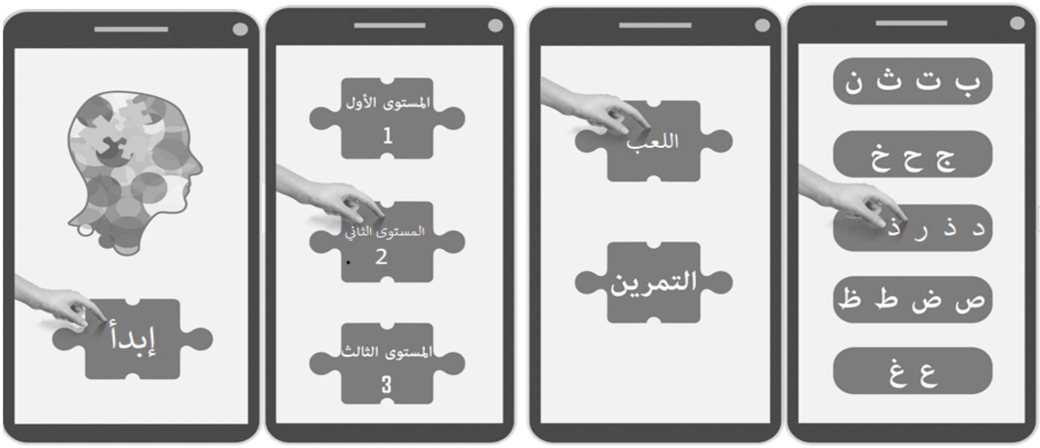


Figure 2. Group of Arabic letters to be identified and read by the participants

After choosing the group of similar notes, the dyslexics used one letter and compared it with more than one vowel, starting from left to right as shown in Figure 3. After choosing the vowel, the app was now ready to start playing and learning. The puzzle image and letters were used for the exercise. It started by asking the children to complete the puzzle. Later, any missing word from the puzzle ended in the final exercise where they had to choose the correct letter to make the word correct.

The puzzling sequence viewing distances were set not far apart. The participants were instructed to make sure that their eyes were kept in focus and followed the puzzle position from the starting point of the trial to the end. Each trial started with the first sequence of the puzzling and ran through the last. Hence, the eye tracking in the focused condition was noticed on the target location of the puzzle piece. Where the participants got the missing puzzle piece right or wrong, their eye movement was recorded, in addition to the time taken for that movement. The participants were also instructed to identify and compare their target within the puzzle.

Figure 3. The puzzled sequence of Arabic letters

3.2. DEMATEL evaluation criteria

DEMATEL is regarded as one of the key methods for evaluating interactive user experiences by linking the various functions in identifying the interdependence among by evaluation (Si et al., 2018). This technique has gained popularity due to its ability to indicate whether each criterion belongs to the cause or the effect group. The technique allows participants to make judgments by assigning weightages to show how one criterion influences another criterion. This allows to establish the interdependent relationships among the factors. As a result, they can be ranked for long term strategic decision. There are two major procedures for undertaking DEMATEL evaluation analysis, namely data gathering/collection step and interrelationships analysis step.

The data collection requires gathering expert opinion. In this study, the parents of the participants that participated in responding to a structural questionnaire that was designed based on the user's experience attributes developed by Morville (2004). The DEMATEL requirement for "expert" is referring to "the awareness of the subject." The justification of adopting Morveille's user experience design factors lie with the fact that apps meant for aiding individual learning processes, especially dyslexics, should contain certain features such as speaking aid, reading aid, spelling aid, and writing aid. Although visual aids were mostly ignored for the reason that dyslexia do not involve visual disability, Stein (2014) argued that "many dyslexics have problems with clearly seeing letters and their order" This is an abnormal condition of visual magnocellular nerve cells that mediate the capability to swiftly recognize letters and their order. It is the central control of visual guidance of individual attention of eye fixations. Based on this observation, the user experience attributes for software programs intended for helping dyslexics in ways that are in line with the evidence on how to help dyslexics are evaluated. According to Morveille (2004), seven important factors are believed to be the user experience criteria, namely: "Useful," "Usable," "Findable," "Credible," "Desirable," "Accessible," and "Valuable." "Useful" refers to the benefits obtained from an eye tracking Arabic alphabetic puzzle app for dyslexia. "Usable" means how an eye tracking Arabic alphabetic puzzle app for dyslexia effectively and efficiently achieves the learning aid for dyslexics. "Findable" refers to how easier it is to find the missing puzzle piece in an eye tracking Arabic alphabetic puzzle app for dyslexia. "Credible" is the ability of the user to trust the learning path established through an eye tracking Arabic alphabetic puzzle app for dyslexia. "Desirable" refers to the extent which dyslexics desire to use an eye tracking Arabic alphabetic puzzle app for dyslexia. "Accessible" refer to how the content of an eye tracking Arabic alphabetic puzzle app for dyslexia reachable to dyslexics. "Valuable" refers to the value an eye tracking Arabic alphabetic puzzle app for dyslexia towards learning. The questionnaire was used to collect data in order to assess the relationships between the established factors. These were decided by the expert and used to establish a group direct-influence matrix. The direct relation matrix indicated a direct influence each factor on the other and evaluated through an integer scale of "no influence (0)," "low influence," "medium influence," "high influence," and "very high influence" as precondition of using this analysis. Therefore, the influence of each criterion is x_{ij}, where $_i$ and $_j$ were assigned as the cause and effect criteria respectively. As a result, the total number of the participants' responses could be represented by $n = 1, 2\ldots, n$, and $n\times n$ was a non-negative direct relation matrix generated by equation 1.

$$x^y = [x_{ij}^y]_{n\times n} \quad (1)$$

where y was the total number of responses by each expert for which $1 \leq y \leq q$ and matrix q for $x^1, x^2, \ldots, x^q$ where q was also the number of responses. The average expert's decision matrix, $Z = [z_{ij}]$ was generated by equation 2:

$$z_{ij} = \frac{1}{l}\sum_{k=1}^{l} z_{ij}^{k}, \quad i, j = 1, 2, \ldots, n. \tag{2}$$

Immediately after generating the direct influence matrix, the next step of normalizing the direct influence matrix to X was taken where X = Z/s, and ***s*** was defined as:

$$s = \max\left(\max_{1\le i\le n}\sum_{j=1}^{n} z_{ij}, \max_{1\le i\le n}\sum_{i=1}^{n} z_{ij}\right) \tag{3}$$

After normalizing the direct influence matrix, the total influence matrix (T) was generated by

$$T = X + X^2 + X^3 + \cdots + X^h = X(I - X)^{-1}, \quad \text{when } h \longrightarrow \infty,$$

where ***I*** was denoted as an identity matrix that generated the Cause and Effect Relationship. The parameters that determined the causal and effect were D and R respectively, where the following equations were used:

$$D_i = \sum\nolimits_{i=1}^{n} t_{ij}, i = 1, 2, \ldots, n \qquad R_i = \sum\nolimits_{j=1}^{m} t_{ij}, j = 1, 2, \ldots, m$$

Here, R_i represented the direct and indirect influence of the criteria ''*i*'' over the other criteria, and D_j represented the influence received by "*j*" by the other criteria. The causal relationship graph was developed by the net effect value and thus ***D-R*** and ***D+R*** were computed. ***D*** and ***R*** represented the sum of the rows and the sum of the columns, respectively, from the total-influence matrix. ***D+R*** indicated how much importance the criterion had and showed the degree of relation between a criterion and all the other criteria. High value of ***D+R*** meant that the criterion had a high relation with other criteria while low ***D+R*** meant that the criterion had a low relation with other criteria. ***D-R***, on the other hand, indicated the kind of relation among criteria. While a positive ***D-R*** indicated that the criterion belonged to the cause group (dispatcher), a negative ***D-R*** indicated that the criterion belonged to the effects group (receiver).

4. Analysis and presentation results

After pre-field experimental study which familiarized the participant on the use of the Aljojo et al. (2018) software program, the recognition, reading and matching of the missing puzzled were set out. Each participant's interaction with three Arabic letter words in the software program for the "Arabic Word Recognition" at phonological level, "Finding missing puzzle" at morphological level, "Fixing missing puzzle" at morphological level, and "Arabic Word Meaning" at lexical level over some period of time was recorded (see Figure 1). The number of errors made during the experiment was collected.

The longest time it took for the Arabic word recognition from the start of the experiment to the end for all participants was found to be seven (7) seconds, as measured in the case of two participants (see Figure 3). Only one participant was able to recognize the three words within two seconds, which is the fastest time it took for the recognition.

The period of finding and fixing the Arabic word puzzle by the participant was recorded (see Figure 4). The fastest time was 43 seconds, whereas the longest time it took to find and fix the missing puzzles was 77 seconds. Finding the puzzle piece was believed to occur after recognition at the beginning of the experiment. The task of finding the puzzle piece involved first understanding the unscrambled Arabic letters and words. The participants picked and entered the missing letters into the puzzle. Therefore, it took some time for the participants to make their decisions, which was the part in which the dyslexics' abilities were measured. As a result, the participants' eye movements within the visual-spatial attention were examined to be directed by the decision after recognition and before finding the missing puzzle piece (see Figure 5).

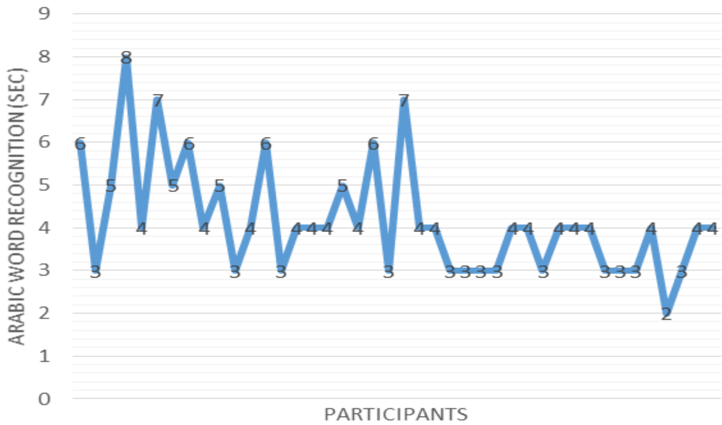


Figure 4. Frequencies of Arabic word recognition for dyslexic children

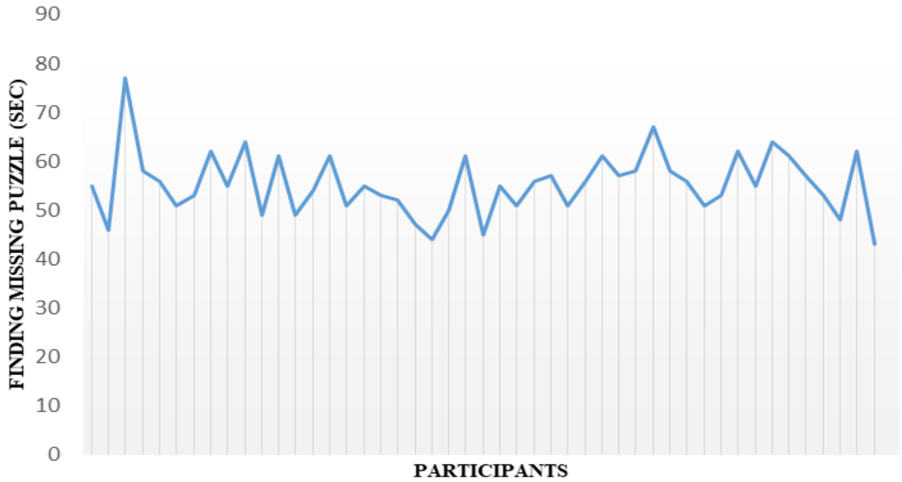


Figure 5. Time taken to find and fix Arabic word puzzle pieces

After Arabic word identification, finding and fixing the missing puzzle piece, the final part was to identify the meaning of the words. The sequence of the events was completed when the participant read the words after finding the missing puzzle and describing its meaning. One of the major results for this last task was to determine whether the participants correctly described the meaning of the word. The time taken for the participant to describe the meaning of the Arabic word was recorded (see Figure 6). The fastest time was two seconds, which was accomplished by most of them while the longest time it took for a participant to describe the meaning was four seconds.

Further analysis of the experimental results revealed that the level of word recognition and understanding of the meaning improved from the analysis of Covariance (ANCOVA). Word recognition and understanding of the meaning are two variables that are measure in order to see the degree of how they joint together, similar to how Green and Bavelier (2003) used covariance measure. Not including the preliminary analysis of Finding and Fixing the Missing Piece, Word Recognition as the dependent variable while the possible effects of Describing the Meaning of Words and Time for Recognition were controlled by entering them as covariates. The Time for Recognition was significant ($F(1,42) = 4.72$, $p = .002$, $\eta^2 = .12$) while the treatment of post-hoc comparisons indicated that the participant's Word Identification significantly decreased their Word Meaning Description. There was an improvement in Finding and Fixing the Missing Piece and the Description of Word Meaning. The time used for the tasks (in seconds) showed an improvement of Description of Word Meaning and the majority of the participants had improved in Word Meaning Description accuracy. The accuracy of the completed Description of Word Meaning over time and Eye Movement tasks was analysed in order to evaluate the visual-spatial attention (decision processing) and regression analysis was carried out. The dependent variable was Arabic Word Meaning Description and the predictors were Finding and Fixing the Missing Piece and Arabic Word Identification. Finding and Fixing the Missing Piece accounted for 71% of the unique variance of being able to describe the meaning of the words ($p = .001$), thus demonstrating that the use of an eye tracking and chatbot Arabic alphabetic puzzle game application for dyslexia did indeed result in reading remediation.

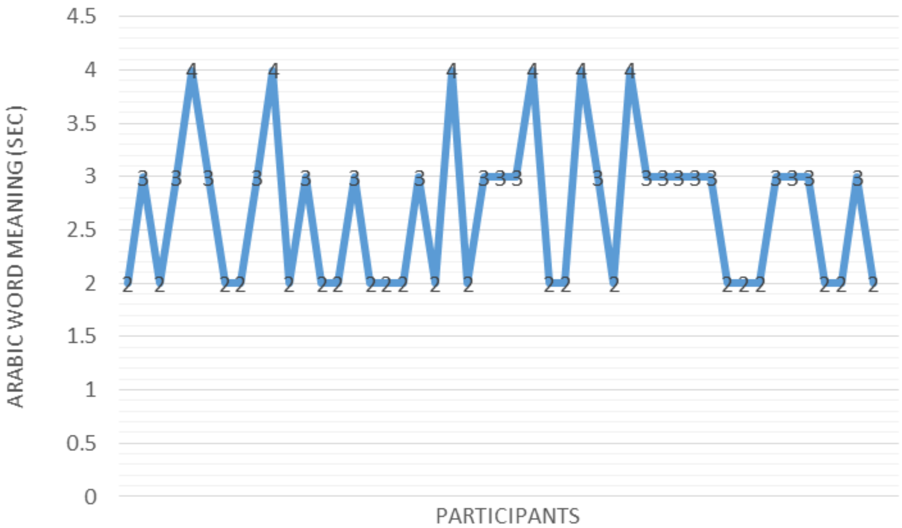


Figure 6. Time taken for revealing the meaning of the Arabic word

The analysis of the key criteria that influenced the app used for dyslexics was carried out using the DEMATEL approach. After the data collection, the first step in obtaining the analysis result involved coding the criteria and entering the data in MS Excel. The criteria were coded as follows: "Useful (UF)," "Usable (UA)," "Findable (FD)," "Credible (CR)," "Desirable (DE)," "Accessible (AC)," and "Valuable (VL)." After calculating the direct influence matrix with equation 1, the result was presented in Table 1. Thereafter, the direct influence matrix was normalized by equation 2 and the result was presented in Table 2.

Table 1. Direct influence matrix

Criteria	UF	UA	FD	CR	DE	AC	VL
UF	0.000	3.207	2.929	2.897	3.069	3.207	2.931
UA	3.267	0.000	3.267	3.172	3.207	3.167	3.067
FD	2.821	2.714	0.000	2.923	2.929	2.929	2.786
CR	2.821	3.000	2.893	0.000	3.154	2.821	2.821
DE	2.857	3.148	3.250	3.321	0.000	3.357	3.107
AC	3.107	2.929	3.036	3.107	2.964	0.000	2.821
VL	3.036	2.929	2.750	2.750	2.821	2.964	0.000

Table 2. Normalized direct influence matrix

Criteria	UF	UA	FD	CR	DE	AC	VL
UF	0.000	0.168	0.153	0.151	0.160	0.168	0.153
UA	0.171	0.000	0.171	0.166	0.168	0.165	0.160
FD	0.147	0.142	0.000	0.153	0.153	0.153	0.146
CR	0.147	0.157	0.151	0.000	0.165	0.147	0.147
DE	0.149	0.164	0.170	0.173	0.000	0.175	0.162
AC	0.162	0.153	0.159	0.162	0.155	0.000	0.147
VL	0.159	0.153	0.144	0.144	0.147	0.155	0.000

The cause and effect relationship matrix was generated and the result presented in Table 3. D and R represent the sum of the rows and the sum of the columns respectively (see Table 3).

The final evaluation involved establishing the relationship among the cause and the effect criteria. According to the rule-of-thumb, positive ***D-R*** indicated that the criterion belonged to the cause group, while a negative ***D-R*** indicated that the criterion belonged to the effects group (see Table 4). Therefore, UF, UA and DE fell within the cause group, which indicated that they influenced other criteria, while FD, CR, AC and VL fell within the effect group, which meant that they were influenced by other criteria.

D+R indicated how much importance the criteria had and showed the degree of relation between a criterion and all the other criteria. A high value of ***D+R*** meant that the criterion had a high relation with other criteria, while a low value meant that the criterion had a low relation with other criteria. Since ***D+R*** determined the importance of a criterion with respect to other criteria, Table 5 indicated the ranking of the enabling criteria according to importance. The table revealed that "Desirable" criteria constituted the most important enabling criteria followed by the "Usable" criteria, while the "Valuable" criteria were the least important.

Table 3. Cause and effect relationship matrix

Criteria	UF	UA	FD	CR	DE	AC	VL	D
UF	-0.149	0.047	0.026	0.023	0.035	0.042	0.031	0.056
UA	0.044	-0.156	0.042	0.035	0.038	0.032	0.033	0.067
FD	0.030	0.022	-0.141	0.035	0.035	0.033	0.030	0.045
CR	0.027	0.038	0.029	-0.145	0.047	0.022	0.029	0.048
DE	0.016	0.037	0.042	0.046	-0.157	0.047	0.037	0.067
AC	0.042	0.030	0.036	0.040	0.030	-0.151	0.026	0.054
VL	0.043	0.036	0.022	0.022	0.027	0.034	-0.138	0.045
R	0.053	0.053	0.055	0.056	0.056	0.059	0.049	

Table 4. Direct influence of the criteria among themselves

Criteria	D	R	D - R	D + R	Criteria type
UF	0.056	0.053	0.002	0.109	Cause
UA	0.067	0.053	0.014	0.121	Cause
FD	0.045	0.055	-0.011	0.100	Effect
CR	0.048	0.056	-0.008	0.103	Effect
DE	0.067	0.056	0.011	0.123	Cause
AC	0.054	0.059	-0.005	0.113	Effect
VL	0.045	0.049	-0.004	0.095	Effect

Table 5. Ranking of enabling criteria

Criteria	D	R	D - R	D + R
DE	0.067	0.056	0.011	0.123
UA	0.067	0.053	0.014	0.121
AC	0.054	0.059	-0.005	0.113
UF	0.056	0.053	0.002	0.109
CR	0.048	0.056	-0.008	0.103
FD	0.045	0.055	-0.011	0.100
VL	0.045	0.049	-0.004	0.095

5. Discussion

This study has investigated the dyslexic sequence of learning using an Arabic software program to complete three major tasks: Word Recognition, Finding the Missing Piece, Fixing the Missing Piece, and finally Arabic Word Meaning. The software program is called "Eye tracking and chatbot Arabic alphabetic puzzle game application for dyslexics." The key criteria involved in the app aiding the learning sequence were also examined. The program has been used with dyslexics to help them alleviate their specific learning disability of reading and spelling manifested in individuals experiencing difficulties in word recognition. Even though certain genetic and environmental traits are associated with it (Peterson & Pennington, 2015), there is no common consensus in regard to its ultimate cause (Benfatto et al., 2016). Generally, a correctional approach is being taken, especially through educational and tutoring software programs. This paper has examined the sequence of learning in Arabic text by Saudi Arabian dyslexics using the software program for aiding dyslexics learning approach. Despite the previous studies that evaluate software programs for dyslexics, they did not focus on specific learning processes.

The findings of the research indicated that the longer it took the dyslexic participants to recognize the Arabic word, the more of an error will be established. In all cases were time was measured from identification, puzzling, and interpreting the meanings of the given Arabic words, it could be observed that the participants managed to complete their task within a few seconds. The engagement with the device when running the application took some time, but the tasks involving Arabic word recognition and fixing the Arabic word puzzle were completed within two minutes. Another important aspect of the result involved the longest time taken to find and fix missing puzzle pieces and the "error" that was generated during the experiment. It could be observed that most errors were caused by the eye movements across the small-sized mobile device screen. Thus, the state of visual-spatial attention affected finding the puzzle piece. The error was not quantified, rather it was observed. Each participant's eye movement was observed in terms of identifying the correct or the wrong puzzle piece. The counted errors were associated with the participants' inability to find the right pieces. That is, the participants were unable to locate the right puzzle while their eye moved around making decisions, yet the wrong ones. The detailed sequence of the participants' experimental events indicated an improvement of reading abilities as the participants' rate of identifying Arabic words and describing their meanings increased after the preliminary experiment. The participants were able to identify Arabic word correctly within the time limit that was less than the time it had taken them before using the app. They were also able to describe Arabic word meanings accurately within a short period after using the app. Overall, the shortest time recorded in describing the meaning of Arabic word was two seconds and achieved by the majority of the participants.

The connection between this research finding and previous work lies with the fact that the potential utility of recording children's eye movements during reading when conducting future applied research is "repeated overt observation" (Rayner, 2011). That is why this study performed such experiments. The finding revealed the frequencies of Arabic word recognition for dyslexic children. The time taken to find and fix Arabic word puzzle pieces has been drawn. The time taken for revealing the meaning of the Arabic word affect eye-movement was recorded. These records are obtained during the reading. Subsequently it was uses to determine the impact of Dyslexics reading capabilities. Arabic word affects eye-movement when reading and subsequently determine the impact of Dyslexics reading capabilities. Considering the experimental finding on Identifying Arabic Words, Using the Puzzle Piece for Matching Arabic Words and Describing the Meaning of the word or sentence formed, major implications of the research with regard to software applications for aiding dyslexics could be found. The central significance of the study involves the serial position effect in any software program for aiding dyslexics. This study has contributed in establishing a sequence of interaction with a software program for dyslexics. The experimental outcome shows an increase in reading skills that allowed the users to remember what had been learnt first and the sequence of the events until the last event that was learnt in a series.

This paper has also investigated the key attribute of the dyslexic app that influences all other criteria responsible for providing appropriate user interaction. Morveille's user experience design factors were used. According to the analysis of the ranking of the enabling criteria based on importance, "Desirable" was the most important group of criteria with the highest score. Morveille (2004) described "desirability" as a factor of user experience design to mean the how much that app displays an "emotional" design to include aesthetics that conveyed through branding identity. The major aspect of "Desirable" criteria is to express how design elements are used within a given app to evoke emotion and appreciation. This finding is directly concerned with visual-spatial attention meaning to direct attention to the location of what should be learnt during reading within the app that is evaluated. Since it is concerned with what the app displays, it is also tied with how users engage and interact with the app. Therefore, visual-spatial attention is found to influence reading abilities (Gabrieli & Norton, 2012).

6. Research limitations and suggestions for future studies

Although this study was designed to address the research problem as accurately as possible, the results and findings should be interpreted in light of the inherent limitations. Specifically, the study relies on Rayner et al. (2013) findings on children's eye movements during reading and the use of user's experience attributes developed by Morville (2004). The used of key-informant procedure involves children and parents. Both participated in the reconnaissance experiment and final experiment. Even though experimental test for ethnographic and DEMATEL method did not indicate the presence of method bias, it is recommended that future research should consider other forms of minimizing the bias through the use of multiple respondents where the source of answering will involve reliability and validity test. One way to overcome this limitation is by using objective measures. Another apparent constrain of this research is the nature of the experimental evaluation. While the findings of this study may be beneficial in understanding the nature of the dyslexics use of software program, the findings may also be validated by a different approach other than ethnographic experiment and DEMATEL evaluation. Thus, future studies may incorporate the use other experimental evaluation to overcome the shortcoming caused by the ethnographic evaluation.

7. Conclusion

This paper presents a study to understand the dyslexic sequence of learning in Arabic text on an Arabic alphabetic puzzle app for dyslexics. In general, dyslexics face difficulties in learning to read and studying with a tutoring app to reduce their difficulties. The nature of learning through special software programs has not been the focus of past research, especially the sequence of learning interaction with those software programs. This study has outlined Arabic Word Recognition, Sequence of Finding and Fixing Puzzle Pieces, and Ability to Describe the Meaning of Words. An experimental evaluation of the interaction with the software program for dyslexics with 42 participants revealed the rate at with these three sequences of learning with and without using the software program for dyslexics. A decision-making analysis experiment was carried out in order to determine the key interaction aspect of user experience on the Arabic alphabetic puzzle app for dyslexics. "Desirability" has been identified as the key group criteria responsible for interaction with Arabic alphabetic puzzle app for dyslexics and its impact on visual-spatial attention and acceleration of word recognition and determination of their meaning for Arabic-speaking dyslexics. Even though there were quite a number of errors during the experiments by the participants. The study also determined the time it takes for the Arabic word recognition and the description of the meaning of words to the highest degree of accuracy. This study has contributed to improving the understanding of the interaction sequence of software programs for providing solutions to the learning problems of dyslexics.

Acknowledgement

This work was supported by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under grant No. (611 - 31 - D1436). The authors, therefore, gratefully acknowledge the DSR technical and financial support.

References

Adlof, S. M., & Hogan, T. P. (2018). Understanding dyslexia in the context of developmental language disorders. *Language, Speech, and Hearing Services in Schools*, *49*(4), 762-773.

Aljojo, N., Munshi, A., Almukadi, W., Hossain, A., Omar, N., Aqel, B., Almhuemli, S., Asirri, F., & Alshamasi, A. (2018). Arabic alphabetic puzzle game using eye tracking and chatbot for dyslexia. *International Journal of Interactive Mobile Technologies, 12*(5), 58-80.

Al-Wabil, A., Zaphiris, P., & Wilson, S. (2007). Web navigation for individuals with dyslexia: An Exploratory study. In C. Stephanidis (Ed.), *Proceedings of Universal Access in Human Computer Interaction. Coping with Diversity: 4th International Conference on Universal Access in Human-Computer Interaction (UAHCI 2007)*, Part I (pp. 593–602). Berlin, Germany: Springer.

Bavelier, D., Green, C. S., & Seidenberg, M. S. (2013). Cognitive development: Gaming your way out of dyslexia? *Current Biology, 23*(7), 282–283.

Benfatto, M. N., Seimyr, G. O., Ygge, J., Pansell, T., Rydberg, A., & Jacobson, C. (2016). Screening for dyslexia using eye tracking during reading. *PloS one*, *9*(11), e0165508. doi:10.1371/journal.pone.0165508

Berget, G., Mulvey, F., & Sandnes, F. E. (2016). Is visual content in textual search interfaces beneficial to dyslexic users? *International Journal of Human-Computer Studies*, *93*(92), 17-29.

Bodis-Wollner, I. (2013). Foveal vision is impaired in Parkinson's disease. *Parkinsonism & Related Disorders*, *19*(1), 1-14.

Carpitella, S., Certa, A., Izquierdo, J., & La Fata, C. M. (2018). A Combined multi-criteria approach to support FMECA analyses: A Real-world case. *Reliability Engineering & System Safety*, *169*(11), 394-402.

Chang, W., Cha, H., & Im, C. (2016). Removing the interdependency between horizontal and vertical eye-movement components in electrooculograms. *Sensors*, *16*(20), 1443-1457.

Chang, B., Chang, C. W. & Wu, C. H. (2011). Fuzzy DEMATEL method for developing supplier selection criteria. *Expert Systems with Applications*, *38*(3), 1850-1858.

Clifton Jr, C., Ferreira, F., Henderson, J. M., Inhoff, A. W., Liversedge, S. P., Reichle, E. D., & Schotter, E. R. (2016). Eye movements in reading and information processing: Keith Rayner's 40-year legacy. *Journal of Memory and Language*, *86*(2), 1-19.

Cockburn, A., Quinn, P., & Gutwin, C. (2015). Examining the peak-end effects of subjective experience. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 357-366). New York, NY: ACM.

Evett, L., & Brown, D. (2005). Text formats and Web design for visually impaired and dyslexic readers: Clear text for all. *Interacting with computers*, *17*(4), 453–472.

Fiedler, N. (2010). Human nervous system and behavioral toxicology. In *Nervous System and Behavioral Toxicology* (pp. 383-398). Auburn, AL: Elsevier Inc.

Franceschini, S., Bertoni, S., Ronconi, L., Molteni, M., Gori, S., & Facoetti, A. (2015). "Shall we play a game?": Improving reading through action video games in developmental dyslexia. *Current Developmental disorders reports*, *2*(4), 318-329.

Franceschini, S., Gori, S., Ruffino, M., Pedrolli, K., & Facoetti, A. (2012). A Causal link between visual spatial attention and reading acquisition. *Current Biology*, *22*(9), 814-819.

Franceschini, S., Gori, S., Ruffino, M., Viola, S., Molteni, M., & Facoetti, A. (2013). Action video games make dyslexic children read better. *Current Biology*, *23*(6), 462-466.

Franceschini, S., Trevisan, P., Ronconi, L., Bertoni, S., Colmar, S., Double, K., & Gori, S. (2017). Action video games improve reading abilities and visual-to-auditory attentional shifting in English-speaking children with dyslexia. *Scientific reports*, *7*(1), 5863-5851.

Frutos-Pascual, M., & Garcia-Zapirain, B. (2015). Assessing visual attention using eye tracking sensors in intelligent cognitive therapies based on serious games. *Sensors*, *15*(5), 11092–11117.

Gabrieli, J. D., & Norton, E. S. (2012). Reading abilities: Importance of visual-spatial attention. *Current Biology*, *22*(9), R298-R299.

Gori, S., Seitz, A. R., Ronconi, L., Franceschini, S., & Facoetti, A. (2016). Multiple causal links between magnocellular-dorsal pathway deficit and developmental dyslexia. *Cerebral Cortex, 26*(11), 4356-4369.

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423*(29), 534–537.

Jasri, D. S., & Rahim, R. (2017). Decision support system best employee assessments with technique for order of preference by similarity to ideal solution. *International Journal of Recent Trends in Engineering & Research*, *3*(3), 6-17.

Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language*, *65*(1), 42-73.

Lee, H. S., Tzeng, G. H., Yeih, W., Wang, Y. J., & Yang, S. C. (2013). Revised DEMATEL: Resolving the infeasibility of DEMATEL. *Applied Mathematical Modelling*, *37*(10-11), 6746-6757.

Lotum, G. (2017). *4 Pics 1 Word – Free*. Retrieved from https://itunes.apple.com/ us/app/4-pics-1-word/id595558452?mt=8

Morville, P. (2004). *User experience design.* Retrieved from https://www.usj.edu.lb/moodle/stephane.bazan/webdesign/morville.pdf

Nagamatsu, L. S., Liu-Ambrose, T. Y., Carolan, P., & Handy, T. C. (2009). Are Impairments in visual-spatial attention a critical factor for increased falls risk in seniors? An Event-related potential study. *Neuropsychologia*, *47*(13), 2749-2755.

Paracha, S., Inuoue, A., & Jehanzeb, S. (2018). Detecting online learners' reading ability via eye-tracking. In *Optimizing Student Engagement in Online Learning Environments* (pp. 163-185). New York, NY: IGI Global.

Peterson, R. L., & Pennington, B. F. (2015). Developmental dyslexia. *Annual Review of Clinical Psychology*, *11*, 283-307.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *The Psychological Bulletin, 124*(3), 373–422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8),1457–1506.

Rayner, K., Ardoin, S. P., & Binder, K. S. (2013). Children's eye movements in reading: A Commentary. *School Psychology Review*, *42*(2), 223-233.

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(2), 514-526.

Reid, L. G., Shaywitz, S. E., & Shaywitz, B. A. (2003). A Definition of dyslexia. *Annals of dyslexia*, *53*(14), 1-14.

Rello, L., & Ballesteros, M. (2015). Detecting readers with dyslexia using machine learning with eye tracking measures. In *Proceedings of the 12th Web for All Conference – W4A'15* (Articleno.16). New York, NY: ACM.

Shah, H. (2012). *1000 Sight Words Superhero*. Retrieved from https://itunes.apple.com/us/app/1000-sight-words-superhero/id449497768?mt=8

Si, S. L., You, X. Y., Liu, H. C., & Zhang, P. (2018). DEMATEL technique: A Systematic review of the state-of-the-art literature on methodologies and applications. *Mathematical Problems in Engineering*, *43*(10), 1101-1113.

Smythe, I. (Ed.) (2010). *Dyslexia in the digital age: Making IT work*. New York, NY: A&C Black.

Stein, J. (2014). Dyslexia: The Role of vision and visual attention. *Current developmental disorders reports*, *1*(4), 267-280.

Williams, P., & Hennig,C., (2015). Effect of web page menu orientation on retrieving information by people with learning disabilities, *Journal of the Association for Information Science and Technology*, *66*(4), 674–683.

Zhan, Z., Zhang, L., Mei, H., & Fong, P. S. (2016). Online learners' reading ability detection based on eye-tracking sensors. *Sensors*, *16*(9), 1457-1468.

Zhou, X., Shi, Y., Deng, X., & Deng, Y. (2017). D-DEMATEL: A New method to identify critical success factors in emergency management. *Safety Science*, *91*, 93-104.

Popescu, E., & Badea, G. (2020). Exploring a Community of Inquiry Supported by a Social Media-Based Learning Environment. *Educational Technology & Society*, *23* (2), 61–76.

Exploring a Community of Inquiry Supported by a Social Media-Based Learning Environment

Elvira Popescu* **and Gabriel Badea**
University of Craiova, Romania // elvira.popescu@edu.ucv.ro // gabriel.badea@edu.ucv.ro
*Corresponding author

(Submitted February 4, 2020; Revised April 6, 2020; Accepted June 21, 2020)

ABSTRACT: The Community of Inquiry (CoI) framework has been widely used in blended and online educational research, more recently being applied also to social media settings. This paper explores the learning community created in such a social media-based educational environment, using an extended version of CoI, which includes four components: cognitive presence, social presence, teaching presence and learning presence. The context of study is a project-based learning scenario implemented in an undergraduate Computer Science course. A quantitative content analysis method is employed, to examine a total of 1712 online contributions (blog posts and tweets), generated by 75 students. Results show that the social media tools provide complementary support to the community of inquiry: the blog is primarily a content space (as cognitive presence is dominant), while Twitter is mostly a discussion space (as social and learning presences are dominant). Teaching presence is barely exhibited by the students, being mainly the preserve of the instructor, while learning presence is quite strong, reflecting students' significant self- and co-regulation behavior.

Keywords: Community of Inquiry, Social media, Project-based learning, Computer-supported collaborative learning, Quantitative content analysis

1. Introduction

The *Community of Inquiry (CoI)* framework was proposed by Garrison et al. (2000) and has been used extensively over the past two decades for exploring the development of online learning communities. These can be characterized in terms of three interdependent components (see https://coi.athabascau.ca/coi-model):

- *Cognitive presence* (the extent to which learners can construct meaning through reflection and discourse)
- *Social presence* (the ability of learners to identify with the community and develop interpersonal relationships by projecting their personal characteristics into the community)
- *Teaching presence* (design, facilitation, and direction of cognitive and social processes to support learning).

The framework is still very popular for blended and online educational research, while underlying technologies and practices are continuously changing (Remesal & Friesen, 2014; Swan, 2019). Thus, CoI was initially introduced for computer conferencing, but subsequently extended to other asynchronous communication spaces between students. More recently, it was applied also to social media settings, such as blog (Angelaina & Jimoyiannis, 2012a; Jimoyiannis & Tsiotakis, 2017; Pifarré et al., 2014), wiki (Eteokleous et al., 2014), Twitter (Sinnappan & Zutshi, 2011), Facebook (Kazanidis et al., 2018; Öztürk, 2015) or SecondLife (Burgess et al., 2010). Indeed, these social media tools have been increasingly adopted in educational contexts, with positive effects on the learning process (Anderson, 2019; Ricoy & Feliz, 2016; Yeh et al., 2019). They can be used to foster communication and collaboration between learners and help create online learning networks, actively engaging students in their learning (Lumby et al., 2014; Kuo et al., 2017). Attributes of social, cognitive and teaching presence could be identified in these social learning spaces, thus proving the applicability of the CoI model (Remesal & Friesen, 2014). Nevertheless, the number of studies is limited, so further research is needed to fully investigate the development of online communities of inquiry supported by social media tools.

Furthermore, the CoI framework itself may also be revisited, adapted and enhanced (Swan & Ice, 2010; Remesal & Friesen, 2014). Thus, Shea and Bidjerano (2010) suggested the extension of the CoI model with an additional theoretical construct labeled "*learning presence*." This refers mainly to learner self-efficacy as well as self- and co-regulation, focusing on the active roles of students in terms of metacognitive, motivational, and behavioral traits. Indeed, various analyses identified learner utterances which could not be reliably coded within the original three CoI presences (e.g., attempts to manage time, divide tasks, set goals, or collaboratively try to understand teacher's instructions). This additional presence is aimed to "contribute to a more thorough account of knowledge construction in technology-mediated environments, expanding the descriptive and explanatory power of the Community of Inquiry framework" (Shea & Bidjerano, 2010, p. 1721).

According to Shea et al. (2012), learning presence is noticeably different from teaching presence; the former refers to forethought and planning, monitoring, and strategy use exhibited by students, while the latter refers to instructional design, facilitation of discourse, and direct instruction, which are mainly (though not exclusively) exhibited by teachers. Learning presence is also clearly distinct from the affective and cohesive dimensions of social presence, as well as from the phases of cognitive presence (i.e., triggering event, exploration, integration, and resolution) (Shea et al., 2014). Further studies showed that learning presence "moderates relationships of the other components within the CoI model" (Shea & Bidjerano, 2012, p. 316), is fostered when students are asked to collaborate more deeply and is the only construct significantly correlated with course grades (Shea et al., 2012). Shea et al. (2013) and Shea et al. (2014) provided additional evidence for the validity of the learning presence component, by combining quantitative content analysis and social network analysis (SNA). The construct was further considered in several studies, such as (Hayes, 2014; Hayes et al., 2015; Kreijns et al., 2014; Traver et al., 2014; Wertz, 2014). A coding scheme for learning presence was also proposed by Shea et al. (2012) and subsequently refined by Shea et al. (2014) and Hayes et al. (2015).

The new construct has been applied so far only for analyzing online course discussions, as reported in the above mentioned studies; a blog platform was used in (Shea et al., 2013), but only as a basic means for hosting a single learning journal entry for each student. Hence, we believe it is worthwhile to use the *extended CoI model* also in the context of social media-based learning spaces. More specifically, in this paper we aim to investigate the community of inquiry created in a social learning environment called eMUSE (Popescu, 2014); the context is a project-based learning scenario implemented in an undergraduate Computer Science course. The novelty of our approach comes from using a blend of social media tools (blog and Twitter), rather than a single tool as in related studies (Angelaina & Jimoyiannis, 2012a; Jimoyiannis & Tsiotakis, 2017; Pifarre et al., 2014; Sinnappan & Zutshi, 2011); this offers the potential for comparative analyses and more in-depth investigations. A longitudinal analysis is also performed, to explore the evolution of the community of inquiry over the course of the semester. While many related works rely on post-hoc information collected through surveys (self-reported perceptions), our study has the advantage of being based on the examination of students' actual discourse (Remesal & Friesen, 2014). A quantitative content analysis method is employed, conducted with the help of an in-house tool, called CollAnnotator (Badea & Popescu, 2017). A total of 1712 utterances contributed by 75 students are analyzed (479 blog posts and 1233 tweets), which is a relatively large sample compared to similar studies.

Overall, our study aims to explore how social media tools (in particular blog and Twitter) can promote collaborative interaction between students in a higher education context. Understanding how blogs and Twitter can support collaborative learning and the creation of a community of inquiry is an important issue for both teachers and researchers (Pifarre et al., 2014). Investigating the cognitive, social, teaching and learning presences occurring during the group learning project may shed some light on the educational affordances of the social media-based learning environment. Indeed, many studies outline the importance of exploring CoI presences in various educational settings (Remesal & Friesen, 2014; Swan, 2019). CoI provides a useful framework for assessing students' learning activities and their involvement in the online community supported by social media tools (Angelaina & Jimoyiannis, 2012a). As these tools become more widely used in education, it appears more important to study their potential to foster the development of a community of students engaged in collaborative learning and social knowledge construction.

More specifically, our research aims to answer questions such as: To what extent do students experience social, cognitive, teaching and learning presence in a project-based learning activity supported by social media tools? What are the frequencies of occurrence of the four presences in the students' blog posts and tweets? What are the differences between blog and Twitter in terms of CoI support? How does the community of inquiry evolve over time? Thus, the main contributions of our paper are on two directions: (i) investigating the use of the *extended CoI model* in a social media-based learning environment (which, to the best of our knowledge, has not been attempted before); (ii) applying a quantitative content analysis on students' contributions on a blend of social media tools (blog and Twitter), enriched with comparative and longitudinal analyses.

The rest of the paper is structured as follows: the next section provides an overview of related work. The study settings and data collection process are described in section 3. The content analysis procedure is detailed in section 4, while the results are reported in section 5. A discussion of the findings and some concluding remarks are included in section 6.

2. Related work

In what follows, we present an overview of studies which explore the use of the CoI framework in social media settings, such as blog, microblog, wiki, social network or virtual world. According to the literature, there are two approaches for the investigation of the CoI presences, which we address in turn: (i) content analysis of students' online messages; (ii) questionnaires for gauging students' perception regarding their learning experience.

2.1. Studies based on quantitative content analysis

We start with several papers that explore the affordances of the blog for supporting a community of inquiry.

One of the first studies belongs to Angelaina and Jimoyiannis (2012a), where the instructional scenario consists in using an educational blog for a project-based learning approach. The study participants are 21 high-school students (15 years old), from two different classes in a Greek school, together with their teacher; they used the blog for 9 weeks, in the context of an informatics curriculum. They published a total of 39 posts and 92 comments, which were extracted from the blog and analyzed by the researchers using CoI model; the unit of analysis was the entire post or comment. Results showed the following distribution of blog entries:

- 95 belong to the *Cognitive presence* (*Triggering event* - 14, *Exploration* - 56, *Integration* - 25)
- 22 belong to the *Social presence* (*Open communication* - 15, *Emotional expression* - 3, *Group cohesion* - 4)
- 14 belong to the *Teaching presence* (*Design and organization* - 1, *Facilitating discourse* - 7, *Direct instruction* - 6), most of them originating from the teacher.

The authors conclude that "project-based blogs can support online learning groups where students are able to share content and ideas, and construct knowledge within a supportive CoI" (Angelaina & Jimoyiannis, 2012a, p. 180).

In a subsequent paper, Angelaina and Jimoyiannis (2012b) extended the analysis by using a combined framework: CoI model in conjunction with the representation of learning mapping (e.g., chain, spoke or network structure). Blogging patterns and students' engagement in blogging activities are investigated. Results suggest that "properly designed blog activities can help students to achieve higher cognitive levels through enhancing their collaboration skills and critical thinking" (Angelaina & Jimoyiannis, 2012b, p. 183).

Pifarre et al. (2014) also investigated the use of blogs for supporting a community of inquiry in secondary education. 15 students from seventh and eighth grades in a Spanish school worked on a science blogging project for four months, discussing topics related to astronomy and space sciences. Each student had to create 6 blog posts solving specific learning activities and also add comments on their own and peers' blogs. A total number of 87 comments were posted by the students and their two instructors, whose content was subsequently analyzed by two coders. First, posts were divided into "meaningful units" (480), which were then coded using CoI scheme. The first 20% of the units were coded by both raters and a good inter-rater reliability measure was obtained; hence, the rest of the units were analyzed by one coder. Results showed the following distribution:

- 130 units belong to the *Cognitive presence*
- 230 units belong to the *Social presence*
- 120 units belong to the *Teaching presence.*

No detailed classification is reported in the paper (i.e., at category level). Authors conclude that "the blog environment afforded the construction of a Community of Inquiry and therefore the creation of an effective online collaborative learning community" (Pifarre et al., 2014, p. 72).

Another study was performed by Jimoyiannis and Tsiotakis (2017), who investigated an educational blogging community created in an undergraduate course entitled "Internet Services and Applications." The course took place at a Greek university and included 48 students split in 10 groups of 4-5 members; each group had to create a blog with valuable content on the topic of "internet safety." A total of 1,214 entries were published on the blogs (200 articles, 15 group pages and 999 comments). Their content was coded by two researchers based on CoI model, using each post as unit of analysis. However, the overall frequency counts for each presence and category were not reported in the paper. Instead, only three representative student groups were analyzed; *cognitive* and *social presences* played a significant role for these groups, while *teaching presence* was less well represented. In addition, the authors applied topic analysis (learning mapping representation of blog topics) and SNA (cohesion analysis, power analysis, role analysis) in order to further explore the learning community.

Twitter's affordances for supporting a community of inquiry have been less explored so far. The study performed by Sinnappan and Zutshi (2011) is a notable exception. Two student cohorts were involved in the study, one from an Australian university and one from an American university. They were enrolled in two similar undergraduate courses on eBusiness. Students were encouraged to use Twitter for exchanging thoughts, ideas and questions on topics like privacy, ethics and censorship, for a period of four weeks. The two instructors collaborated in order to foster interaction between students across universities. A total of 47 learners participated in the study, posting 324 curriculum-related tweets. Their content was analyzed by two coders, using CoI scheme. Up to two categories could be assigned to each tweet (unit of analysis): a primary one (mandatory) and a secondary one (optional); 186 tweets had 2 assigned categories. The following distribution of tweets was obtained (taking into account only the first category / both categories):

- 279 / 284 belong to the *Cognitive presence* (*Triggering event* - 82 / 84, *Exploration* - 194 / 197, *Integration* - 3 / 3)
- 13 / 194 belong to the *Social presence* (*Open communication* - 1 / 4, *Affective* - 4 / 8, *Group cohesion* - 8 / 182)
- 32 / 32 belong to the *Teaching presence* (*Design and organization* - 8 / 8, *Facilitating discourse* - 22 / 22, *Direct instruction* - 2 / 2), most of them originating from the teacher.

The study shows Twitter's potential for pedagogical use, being able to enhance and complement all CoI presences (Sinnappan & Zutshi, 2011).

An alternative approach to the manual content analysis performed in the above studies is the use of machine learning and text mining techniques for classifying students' posts (Wu et al., 2020; Xing & Gao, 2018). In recent years, these have started to be used also in the context of the CoI framework, providing automatic labeling of learner messages according to the categories of CoI presences. Such studies have been performed based on the coding scheme for the *cognitive presence* (Farrow et al., 2019; Neto et al., 2018) or the *social presence* (Ferreira et al., 2020). To the best of our knowledge, no automatic methods have been devised so far for classifying messages with respect to *teaching presence* or to the whole CoI model. While this is a promising research direction, it is outside the scope of this paper, which focuses on manual content analysis; furthermore, manual coding can provide a higher accuracy than currently existing automatic approaches.

2.2. Studies based on students' perceptions

CoI model has also been applied to other social media tools, such as wiki, Facebook and Second Life; however, no content analysis has been performed in these cases; instead, researchers relied on the *Community of Inquiry Survey* (Arbaugh et al., 2008) in order to extract students' subjective perceptions regarding the three presences.

Thus, Öztürk (2015) investigated the suitability of a social networking site for supporting a community of inquiry. Facebook was used in a blended course on Philosophy of Education in a Cypriot university. 198 students were involved in the study (from two different cohorts); one large group (77 students) and 5 smaller groups (around 24 students) were formed, who participated in various discussion activities around short videos and articles. Data was collected through CoI Survey and a Motivation Scale, which were filled in by 158 students. Results showed that *teaching presence* perception was highest, followed by *cognitive presence* and then *social presence*; a high correlation among the three presences was determined. Furthermore, all presence perceptions were higher for students in the small groups compared to the large group. The study also found that students with a higher cognitive presence perception have higher academic success, while students with higher teaching presence and cognitive presence perceptions have higher motivation. The authors conclude that Facebook can be effectively used for educational purposes, facilitating the creation of communities of inquiry.

The same social networking site was used by Kazanidis et al. (2018), in an attempt to compare students' learning experience with Moodle and Facebook in a course on Instructional and Learning Theories at a Greek university. 97 students participated in the study, being split into two groups based on the learning and communication platform adopted: 47 learners used Moodle (control group) and 50 learners used Facebook (experimental group). Students worked in teams of 4-5 people over the course of six weeks in order to create various educational resources and reports. At the end of the study, students filled in a revised version of the CoI Survey. Results indicate that students who used Facebook had a higher *social presence* perception compared with students who used Moodle, while *teaching* and *cognitive presence* perceptions were similar for both groups. Furthermore, female students in the experimental group had higher teaching and cognitive presence perceptions compared with their male peers. Overall, the study outlines Facebook's potential to support teaching and learning processes, increasing students' engagement and learning satisfaction.

Eteokleous et al. (2014) evaluated the integration of wiki as educational tool in an elementary school in Cyprus. 20 fifth-grade students were enrolled in a Language and Linguistics class and used a wiki for the course of five lessons, under the guidance of their teacher. Data was collected through several methods: the CoI Survey, reflective journal for the teacher, based on the Dynamic Model of Educational Effectiveness (Creemers & Kyriakides, 2010), observations performed by an external coder, and interviews with students. Results of CoI survey suggest that *teaching presence* was perceived as strongest, highlighting the important role of the teacher, especially in terms of *direct instruction* and *facilitation*. *Cognitive presence* was less strong, but *integration* and *resolution* phases of learning were well supported, while *triggering event* had a satisfactory appearance. *Social presence* was perceived as relatively weak, indicating that students did not have enough opportunities for *open communication* and fostering *group cohesion*; instead, they interacted more offline than through the wiki. Overall, the use of wiki promoted the development of a CoI to a satisfactory degree, contributing to the achievement of educational goals.

Burgess et al. (2010) explored the use of a Multi-User Virtual Environment (MUVE), Second Life, for supporting a community of inquiry. Ten graduate students from a US university (pre-service or in-service educators) were enrolled in an online instructional technology course; the instructor used Second Life for two class meetings. Two instruments were used for gathering data: CoI Survey for students' perceptual data and the Multi-User Virtual Environment Education Evaluation Tool (McKerlich & Anderson, 2008) for observational data, recorded by two coders. High perception levels were reported by the students for all three presences; also, a medium number of observations were recorded by the coders for each presence. Hence, a community of inquiry may be developed inside a MUVE, and the main elements of the CoI model apply to immersive environments, as previously suggested by McKerlich and Anderson (2008).

Finally, the Community of Inquiry Survey was also used in the context of a blogging activity, as reported by Yang et al. (2016). The study included 26 graduate students at a Taiwanese university, who took a Digital Learning course; learners had to write at least two posts per lecture and could also comment and rate their peers' posts. At the end of the semester, students were asked to fill in a questionnaire for identifying their perceptions of CoI presences, as well as their subjective learning outcomes. Regression analysis was applied and results showed that all three presences have a significant role in predicting learning performance: *cognitive presence* played the most important role, followed by *social presence* and to a lesser extent by *teaching presence*.

The works presented above vary in terms of context of study, social media tool, discipline, participants, analysis methods and results. The scale of the studies is relatively small: most of them involve 10 to 50 participants, with the exception of (Öztürk, 2015), which includes almost 200 students. The set of disciplines being taught in the studies is varied, ranging from informatics to education and instructional technology, and from eBusiness to language and linguistics, showing that communities of inquiry may be developed around any topic. While the blog is the most popular medium in existing studies, research has been performed also on the affordances of other social media tools for supporting a community of inquiry, e.g., microblog (Sinnappan & Zutshi, 2011), wiki (Eteokleous et al., 2014), social network (Kazanidis et al., 2018; Öztürk, 2015) or virtual world (Burgess et al., 2010). As far as the analysis method is concerned, most existing studies rely on students' perceptions, gauged by means of the CoI survey, but there are also a few which perform content analysis (Angelaina & Jimoyiannis, 2012a; Jimoyiannis & Tsiotakis, 2017; Pifarre et al., 2014; Sinnappan & Zutshi, 2011). While presence distribution varies from study to study, all surveyed works confirm the suitability of the respective social media tool to support a community of inquiry, to various extents. However, none of them uses the extended CoI model; to the best of our knowledge, the current study is the first one to employ this extended model in social media-based learning settings. Furthermore, all the papers reviewed in this section investigate a single social media tool, while the current study explores and compares a conjunction of tools, as described in subsequent sections.

3. Study settings and data collection

Our study took place in the context of a semester-long (i.e., 14 weeks) course on Web Applications Design (WAD), taught to 4th year undergraduate students in Computer Science from the University of Craiova, Romania. Students followed a collaborative project-based learning (PBL) scenario, in which they had to design and implement a relatively complex web application of their choice (e.g., a virtual store, an online travel agency, an auction website, an educational social network, an online library). A total of 75 students were involved in the study, 18 female and 57 male (average age 22). They were grouped in 20 teams of 3-4 peers, formed according to students' preferences. Each team member took various roles throughout the semester: system analyst, database specialist, interface designer, application architect, programmer, tester, project manager etc. At this stage, students had already taken several programming courses, as well as a Database Design, a Software Engineering and a Project Management course; therefore they had enough knowledge and experience to undertake a team-based development of a real-life software product. At the end of the semester, students had to make a presentation of their product in front of the whole class; in addition, they were asked to give four intermediary presentations, in order to show the progress of the project. The evaluation was based both on the final product and the collaborative work carried throughout the semester.

A blended learning approach was implemented: every week students participated in face-to-face classes with the instructor and used a mix of social media tools for online communication and collaboration. As PBL has a strong social component, being rooted in constructivist principles (Savery & Duffy, 1995), the social media tools can and have been used to support communication and collaboration in the PBL framework (Ardaiz-Villanueva et al., 2011; Kim et al., 2011). More specifically, each team had a dedicated wiki space for collaborative writing tasks, for gathering and organizing educational resources and for documenting each stage of the project. They also had a team blog, for reporting the progress of the project (i.e., learning diary), reflecting on their learning experience, publishing thoughts, ideas, and resources, describing problems encountered and asking for help, providing feedback and solutions. In addition, each student had an individual Twitter account, for staying connected with peers and posting short news, announcements, questions, status updates regarding the project. All these social media tools were integrated in a social learning environment called eMUSE (Popescu, 2014). The platform provided various functionalities, both for students and instructors: easy access to the tools and latest activity notifications, learner tracking and data visualizations, peer review module, grading support. More details regarding the platform can be found in (Popescu, 2014; Popescu & Petrosanu, 2016).

Students had no prior experience with eMUSE platform; however, most of them had used the social media tools before, in out-of-school contexts. The use of the tools was mandatory, as students' contributions were part of their final grade; this was a way of assessing students' collaborative work throughout the semester. More specifically, the intermediary presentations and continuous collaborative work counted for 70% of the grade, while the final project counted for 30%. The instructor provided brief guidelines regarding the use and expected role of each tool at the beginning of the study (as mentioned above); continuous feedback and clarifications were provided throughout the semester. However, no specific scaffolds or prompts were included, so students had a high degree of freedom and flexibility. Thus, every week there was a two hours face-to-face class with the instructor, in which students received hands-on tutorials and help with their projects; subsequently, students used the social media tools in eMUSE for online communication and collaboration when developing their projects at home (every week, after class). The specific amount of activity performed on each tool varied from one week to the other, as detailed in section 5.3.

The aim of our study was to investigate the community of inquiry created around the WAD project and the affordances of the social media tools to support it. More specifically, we were interested to apply the extended CoI model to the communication space created through blog and Twitter. The wiki was not included in this analysis, as it was used especially for writing the project documentation and students did not use the associated discussion pages for further communication. Hence, blog posts, comments and tweets supported all message exchanges between students and they were automatically collected and stored by eMUSE, as mentioned above. Thus, a total of 399 blog posts, 80 blog comments and 1,233 tweets were recorded throughout the semester. Their content was subsequently analyzed, according to the procedure described in the following section.

4. Content analysis procedure

4.1. CollAnnotator tool

We used quantitative content analysis based on the extended CoI model, as presented in the Introduction. This is a popular approach in technology enhanced communication and learning for generating categorizations and frequency counts based on various coding schemes (Hayes et al., 2015).

Indeed, transcript / content analysis can offer significant insights to understand students' interaction patterns and discourse quality in online communities of inquiry (Garrison et al., 2006). However, content analysis is generally a laborious task, therefore an analysis support tool could facilitate the coding and negotiation process (Garrison et al., 2010). A potential solution would be to use generic commercial software for content analysis (e.g., ATLAS.ti (see http://atlasti.com), NVivo (see http://www.qsrinternational.com/nvivo-product), Dedoose (see http://www.dedoose.com) etc.). However, this is costly, more difficult to learn and use, not accommodating CoI specificities, it requires input data in a particular format and does not always offer support for multiple coders.

Hence, a dedicated tool for supporting content annotation based on CoI would prove useful to the researchers. In particular, based on literature reports (Garrison et al., 2006; Garrison et al., 2010; Sinnappan & Zutshi, 2011), we extracted the following set of essential functionalities which need to be provided by such a tool:

- Intuitive and easy to use interface, which requires virtually no learning curve for the coders (i.e., persons who annotate and assign codes / categories to content)
- Rich annotation support, which can be done both at message level (i.e., unit of analysis) but also at higher levels of granularity (e.g., word, phrase, sentence)
- Possibility to attach more than one code to a message (e.g., include a primary and a secondary category)
- Support for multiple coders and suggestive comparisons between them, which may increase the rigor and reliability of the coding process
- Support for the negotiation phase, in which researchers discuss their individual codes and aim to bring them into alignment with each other, pursuing a shared identification of meaning; visualizing the other coders' notes, comments and highlights can substantially aid this process
- Detailed statistics and reports of the coding results, including graphical visualizations.
- Starting from these requirements, we developed a dedicated content analysis tool called *CollAnnotator*. The platform was briefly introduced in (Badea & Popescu, 2017) and a preliminary analysis was reported in (Popescu & Badea, 2017).

In what follows, we provide a concise description of CollAnnotator platform and its functionalities. The tool is adapted to our goal of using CoI for investigating the online community formed in our social media-based learning environment, eMUSE; it directly retrieves student content (blog posts and tweets) from eMUSE database and generates reports and statistics specific to our instructional scenario. The main features provided by CollAnnotator include:

(1) View, annotate and categorize student contributions

The coders can visualize students' blog posts and tweets (in the original HTML format used by the student), as well as search and order them by author, team, date and title. They can subsequently use the extended CoI scheme to assign a primary and an optional secondary category to each post. An explanatory comment may also be added, such as specifying the indicator used for the particular category (e.g., Goal setting, Planning or Coordinating, delegating tasks to self and others for the Forethought & planning category; Expressing emotions, Use of humor, Self-disclosure, Use of unconventional expressions to express emotion, Expressing value for the Affective category (Shea et al. (2014)).

The unit of analysis is the blog post / tweet, based on the recommendations provided by several authors (Angelaina & Jimoyiannis, 2012a; Garrison et al., 2006; Sinnappan & Zutshi, 2011). The coder may also refer to a specific part of the post (i.e., word, sentence, paragraph), by using the highlight, comment and tag functionality provided by CollAnnotator; further justifications for the category selection, as well as a more detailed personal interpretation of the post may be included this way.

(2) Compare and negotiate assigned categories

CollAnnotator offers support for the negotiation phase, which may thus take place online, without the need for a face-to-face meeting between the coders. Each coder can view the categorization chosen by the other(s), as well as their comments, highlights and tags and may choose to change their initial category selection.

(3) Visualize reports and statistics

The researcher can use CollAnnotator to visualize the coding results in various formats: summarizing tables with frequency counts for each presence and category, percentage agreement between coders and Cohen's kappa coefficient, suggestive graphical visualizations of the distribution of presences and categories for each social media tool, over time, as well as at student and team level. Some of these visualizations may be seen in the following section (Figures 1-5).

4.2. Coding procedure

Content analysis was performed by two independent coders, in order to increase the reliability and validity of the results. They used CollAnnotator tool and the coding scheme proposed in (Shea et al., 2010; Hayes et al., 2015).

Each of the 479 blog entries and 1233 tweets were assigned to at least one category; due to the richness of some posts, the coders had the option of selecting also a secondary category. They also added justifications of their choice (e.g., the corresponding category indicator) and provided annotations at finer levels of granularity, using the highlight feature included in CollAnnotator.

Coding took place in two phases: first, all posts were annotated and classified independently by each coder, obtaining an agreement percentage of 85.28% (Cohen's kappa value: 0.818). Secondly, negotiation and discussion took place, and consensus was reached in 98.42% of the cases (Cohen's kappa value: 0.98). A detailed analysis of the coding results is included in the following section.

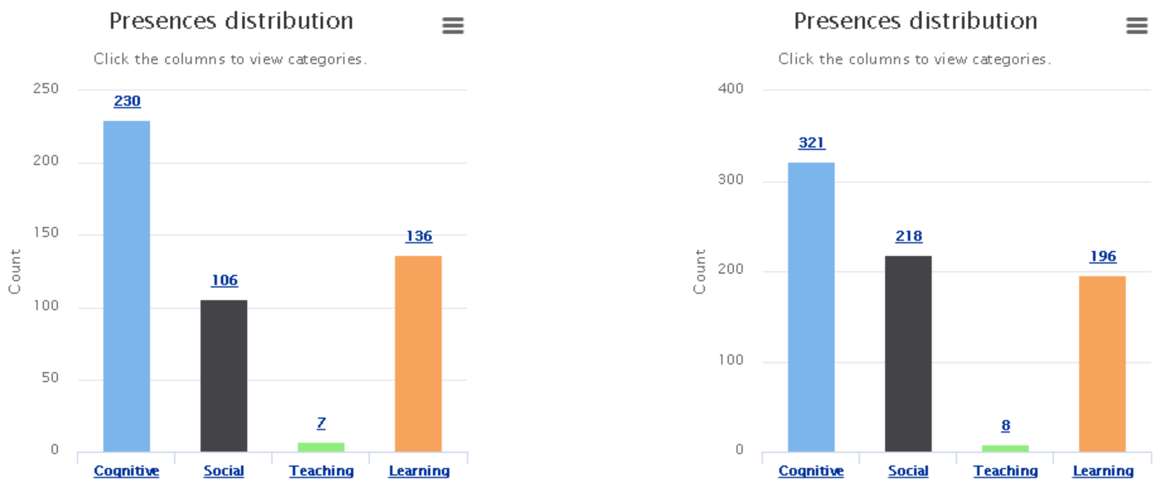
5. Data analysis results

5.1. Blog content analysis according to CoI model

We start by reporting the frequency of occurrence of each presence in the students' blog posts, as computed by CollAnnotator. The graphical distribution of the posts according to the four presences is depicted in Figure 1 and a detailed classification at category level is included in Figure 2.

First of all, we can notice that a large number of blog posts received both a primary and a secondary category (264 out of 479, or more than 55%). This can be explained by the fact that many posts include more than one idea, sometimes belonging to different categories; furthermore, the social component was present in many posts, with a secondary role, as discussed later.

In what follows, we analyze the blog content according to the primary category it belongs to (see Figure 1(a) and Figure 2). Thus, the largest number of posts (230 out of 479, or 48%) belong to the *cognitive presence*. More than half of these (almost 59%) refer to the *integration* phase of learning, as students regularly report on the solutions they created and present various connected ideas and syntheses. The *exploration* phase is also relatively well represented (almost 36% of the posts), with many students sharing interesting resources and ideas (information exchange) or providing suggestions for consideration. On the other hand, the *triggering* phase is scarcely represented (less than 6% of the posts), as students tend to post not when they encounter a problem or puzzlement, but rather when they have a solution or idea to share. Finally, the *resolution* phase is not documented on the blog; this is understandable, as complete, fully-fledged solutions are generally presented and defended on the wiki.



(a) (b)

Figure 1. Number of blog posts pertaining to each presence (*Note:* The graphical representation is generated by CollAnnotator; the left chart (a) takes into account only the primary category associated to a post, while the right chart (b) takes into account both the primary and the secondary category, when available.)

The *learning presence* accounts for more than 28% of the blog posts. Most of these (over 85%) refer to the *monitoring* category, as students frequently report their progress and note the completion of tasks. The rest belong mainly to the *forethought & planning* category (almost 13%), with students setting goals, making plans and assigning project tasks; *strategy use* and *reflection* are scarcely present, as these do not seem to be explicitly verbalized by the students.

Blog Statistics (based on both categories)

Classification Table for Blog Posts & Comments

Presence	Category	Count
Cognitive	Triggering event	17 (13 P + 4 S)
	Exploration	95 (82 P + 13 S)
	Integration	209 (135 P + 74 S)
	Resolution	0
Social	Affective	41 (13 P + 28 S)
	Open communication	89 (42 P + 47 S)
	Group cohesion	88 (51 P + 37 S)
Teaching	Design and organization	6 (5 P + 1 S)
	Facilitating discourse	0
	Direct instruction	2 (2 P + 0 S)
	Assessment	0
Learning	Forethought & planning	44 (17 P + 27 S)
	Monitoring	142 (116 P + 26 S)
	Strategy use	7 (2 P + 5 S)
	Reflection	3 (1 P + 2 S)

Figure 2. Blog summary report (*Note:* The tabular view is provided by CollAnnotator; numbers in parentheses refer to the primary category (P) and the secondary category (S) respectively.)

Social presence is also visible on the blog, in over 22% of the posts. Most of them belong to the *group cohesion* category (over 48%), with many students addressing or referring to the group in their posts. *Open communication* is also well represented (almost 40%), including answers to peers' posts, asking questions, complimenting and expressing appreciation. Only few posts (12%) express emotions or use humor (*affective* category), as students prefer to use the blog in a slightly more formal manner.

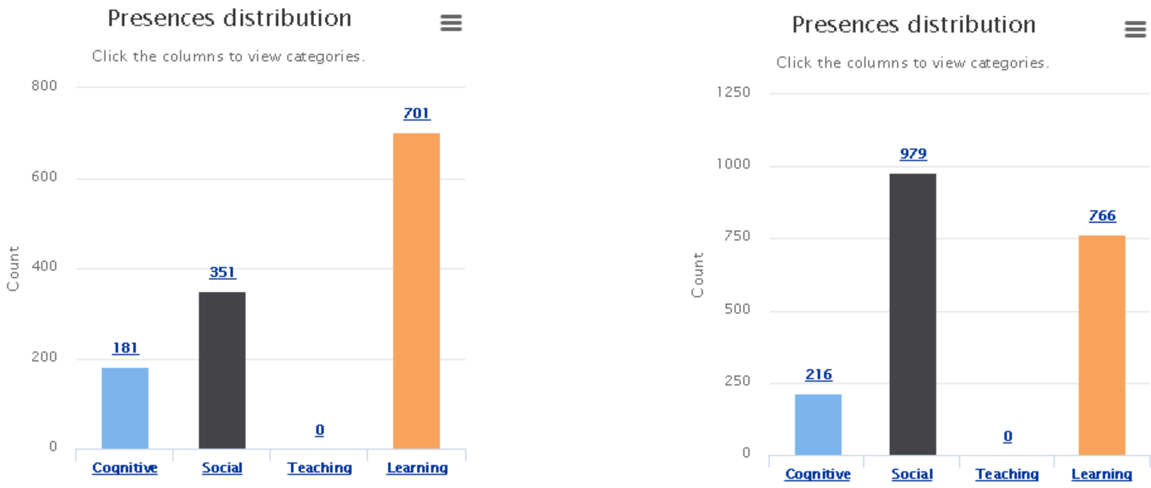
Finally, the *teaching presence* is scarcely exhibited by the students, in less than 2% of the blog posts. However, this is counterbalanced by the instructor's blog posts (that were not included in this analysis), which clarify the design and organization of the course, provide direct instruction and educational material, offer feedback and formative assessment.

When the secondary category is also taken into account (Figure 1(b)), a sharp increase is seen in the *social presence* posts. This can be explained by the fact that many blog entries with a dominant cognitive or learning component also include some social aspects (group reference, salutations and greetings, asking or answering peers' questions). While these do not represent the main focus of the post, they play an important role for strengthening group cohesion and communication. A special mention should be made regarding the *integration* category of the *cognitive presence*; we considered it secondary whenever the student would simply point towards the solution described elsewhere (i.e., on the wiki, in the project documentation), rather than explicitly presenting it within the post. Hence, the number of blog entries belonging to the integration phase of learning significantly increased when taking into account also the secondary category.

5.2. Twitter content analysis according to CoI model

As far as Twitter is concerned, the graphical distribution of tweets according to the four presences is depicted in Figure 3, while a detailed classification at category level is included in Figure 4.

Just as in case of blog posts, a large number of tweets received both a primary and a secondary category (728 out of 1233, i.e., 59%). The secondary purpose of the tweets is in most cases (over 86%) a social one, which is in line with the nature of the medium; students use Twitter for strengthening the group cohesion, by directly addressing their peers (using the *mention* functionality), sending salutations and greetings and sharing various information unrelated to the course.



(a) (b)

Figure 3. Number of tweets pertaining to each presence (*Note:* The graphical representation is generated by CollAnnotator; the left chart (a) takes into account only the primary category associated to a tweet, while the right chart (b) takes into account both the primary and the secondary category, when available.)

Twitter Statistics (based on both categories)

Classification Table for Tweets

Presence	Category	Count
Cognitive	Triggering event	17 (14 P + 3 S)
	Exploration	79 (72 P + 7 S)
	Integration	114 (89 P + 25 S)
	Resolution	6 (6 P + 0 S)
Social	Affective	111 (82 P + 29 S)
	Open communication	194 (176 P + 18 S)
	Group cohesion	674 (93 P + 581 S)
Teaching	Design and organization	0
	Facilitating discourse	0
	Direct instruction	0
	Assessment	0
Learning	Forethought & planning	128 (114 P + 14 S)
	Monitoring	603 (555 P + 48 S)
	Strategy use	32 (29 P + 3 S)
	Reflection	3 (3 P + 0 S)

Figure 4. Twitter summary report (*Note:* The tabular view is provided by CollAnnotator; numbers in parentheses refer to the primary category (P) and the secondary category (S) respectively.)

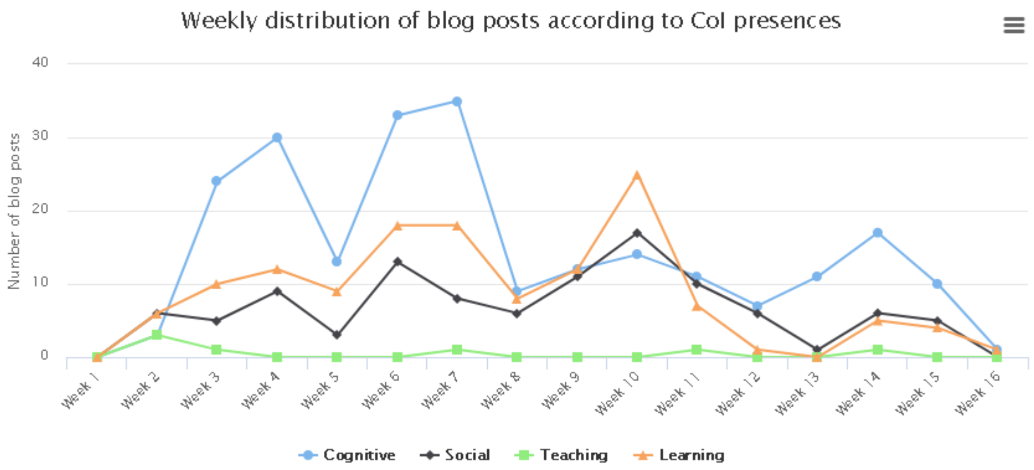
In what follows, we analyze the tweet content according to the primary category it belongs to (Figure 3(a) and Figure 4). Thus, the largest number of tweets belong to the *learning presence* (almost 57%), especially the *monitoring* category; students tweet whenever they complete a task, but also when they set a goal, make a plan or distribute tasks to peers.

The *social presence* is also well represented on Twitter, being the primary focus of over 28% of the tweets. *Open communication* category is supported in more than 50% of them, with students using the *reply* functionality frequently, and also asking questions, expressing agreement or disagreement. *Group cohesion* and *affective* categories are also present in the tweets; students directly refer to the group, use humor and express emotions in unconventional ways.

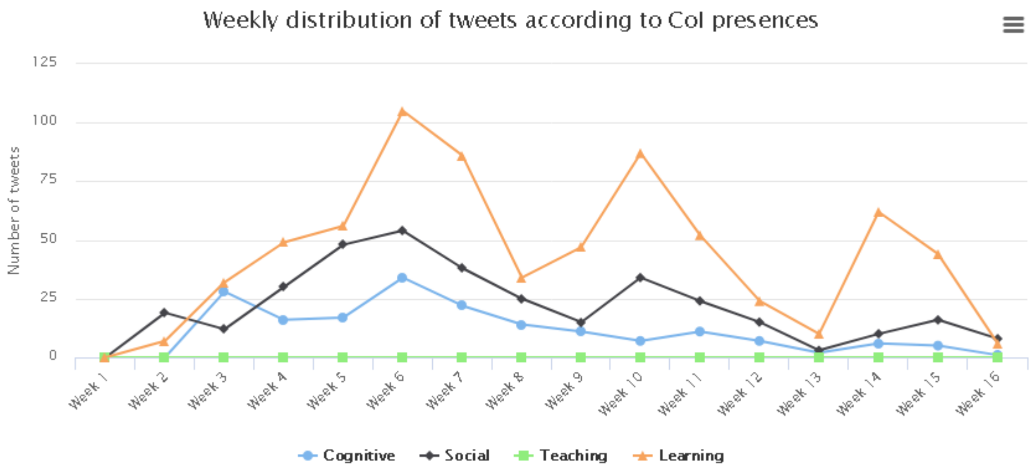
The *cognitive presence* is only apparent in less than 15% of the tweets; students mostly point to created solutions (*integration* category) and share useful educational resources (*exploration* category). Finally, the *teaching presence* is not supported on Twitter, as students do not get involved in any direct instruction, course design or assessment tasks.

5.3. CoI presences distribution throughout the semester

CollAnnotator also provides us with the temporal evolution of students' contributions on blog and Twitter, as illustrated in Figure 5. First of all, we can notice that the weekly amount of posts follows a somewhat similar pattern both on blog and Twitter. Furthermore, within each tool, the weekly patterns of each presence are relatively similar, showing a balanced distribution throughout the semester. As far as the blog is concerned, we notice that the *cognitive presence* is the dominant one in most of the weeks, as expected. Similarly, the *learning presence* is the dominant one on Twitter almost every week.



(a)



(b)

Figure 5. Temporal distribution of students' blog posts (a) and tweets (b) (Note: The graphical representation is generated by CollAnnotator; the four presences are based on the primary category associated to each student contribution.)

Overall, we can see that students used the social media tools throughout the semester, contributing in a continuous manner to the formation and growth of the community of inquiry. Generally, students have the tendency to postpone most of the work for the end of the semester, just before the due deadline (Zacks & Hen, 2018). Our aim was to discourage academic procrastination by means of the instructional scenario proposed, as described in section 3. Thus, the four intermediary milestone presentations encouraged students to work in a sustained manner throughout the semester. Of course, there are some variations over time, as students worked from home, at their own pace; furthermore, learners were enrolled in five other courses, each with their specific time requirements. However, there was some activity recorded every week of the semester, even during the winter holidays (weeks 12 and 13), indicating students' continuous engagement in the community of inquiry.

To sum up, the overall distribution of presences is different on the two social media tools. Thus, when taking into account only the primary category, the *cognitive presence* is dominant on the blog (accounting for 48% of the blog posts), being less apparent on Twitter (accounting for less than 15% of the tweets). Conversely, the *learning presence* is dominant on Twitter (accounting for almost 57% of the tweets), and less represented on the blog (accounting for just over 28% of the blog posts). When considering also the secondary category, the *social presence* becomes dominant on Twitter (almost 50% of the codes), while being less prominent on the blog (just over 29% of the codes). Finally, the *teaching presence* is scarcely exhibited by the students on both social media tools (accounting for less than 2% of the contributions).

6. Discussion and conclusions

According to the content analysis, the blog plays the primary role of a content space, as the cognitive components clearly outweigh the social and learning components. Twitter, on the other hand, is mostly a discussion space, supporting especially the learning presence and the social presence, while the cognitive presence is less strong. Hence, we can conclude that each tool fulfills its own distinct role in the learning space, providing complementary support to the community of inquiry.

The blog results are in line with the findings in (Angelaina & Jimoyiannis, 2012a; Jimoyiannis & Tsiotakis, 2017; Yang et al., 2016), all indicating that cognitive presence plays the most important role, followed by social presence and lastly by teaching presence. The results obtained by Piffare et al. (2014) are different, stating that social presence is the dominant one. This finding can be explained by the fact that only blog comments were analyzed in that study, not also the original blog posts, which likely contained more cognitive elements.

Regarding Twitter, our results are significantly different than those reported by Sinnappan and Zutshi (2011), who found a very strong cognitive presence in students' tweets. Two potential explanations arise: first, only tweets explicitly related to the course (i.e., containing the course hashtag or replying to one of the peers) were included in that analysis; according to the authors, other discussions took place during the study period, which were not related to the scheduled teaching activities. It is likely that these additional tweets had a more social nature, including sharing of information unrelated to the course, personal details, emotions and humor, or communication with purely social function. Secondly, the instructional scenario reported in (Sinnappan & Zutshi, 2011) includes only Twitter as communication tool; so, all questions and problems, suggestions and opinions, information exchanges, solutions and syntheses had to be shared by means of tweets. In our scenario, students could choose the blog for sharing these cognitive issues, while using Twitter more for the social aspects.

A special mention should be made regarding the teaching presence. This was very weak in our study, which is in line with findings from (Angelaina & Jimoyiannis, 2012a; Sinnappan & Zutshi, 2011; Shea et al., 2014). Furthermore, most of the teaching-related posts in these studies originated from the instructor. In our case, teacher's posts were not considered in the analysis, hence the even lower percentage of teaching presence recorded. Nevertheless, the instructor played an important role, as guide, facilitator and observer throughout the semester; her posts helped clarify the design and organization of the course, provided direct instruction, educational materials, feedback and assessment.

The learning presence, on the other hand, was apparent in numerous blog posts and tweets. This supports the findings of Shea et al. (2012), who ascertain that learning presence is more apparent when students are asked to actively collaborate through instructional design. Indeed, students exhibited significant self-regulation and effort regulation behavior, monitoring their progress, setting goals and distributing project tasks. Hence, the decision to integrate this complementary construct in our coding scheme appears well justified. Further research on learning presence in social media-based educational environments could shed more light on its role and relationship with the other presences.

In addition, further studies on the role played by the social media tools in the development of communities of inquiry are welcome. Different learning scenarios, educational tasks and instructor guidelines may influence student participation and distribution of presences, independently of the employed tools. In the current PBL scenario, students had a high degree of freedom and flexibility in using the social media tools for their communication and collaboration activities. While brief guidelines regarding the use and expected role of each tool were provided at the beginning of the semester, no additional scaffolds or prompts were included. Nevertheless, the PBL tasks were complex, challenging and authentic and students worked relatively autonomously, with the teacher playing the role of facilitator. They had to collaborate in various design, problem-solving, decision making and investigative activities, which may explain the dominance of the cognitive and learning presence. Furthermore, the scoring took into account all students' contributions on the social media tools, as a way of assessing collaborative work throughout the semester; this may have been another factor boosting students' engagement in the community of inquiry.

Overall, our study showed that students' contributions in the social media-based learning environment can be characterized in terms of the four extended CoI presences. It also illustrated the potential of the proposed environment to support students in creating an online community conducive of self-regulated collaborative learning. Thus, on one hand, the paper adds to the limited research literature on the extended CoI model, proving its applicability in social media-based learning settings. On the other hand, the current research contributes to a better understanding of students' collaborative processes in an online learning community.

A potential limitation of our study is that the collected and analyzed data was restricted to blog posts and tweets. Of course, students used also other private communication channels, like email, chat, phone calls or face-to-face meetings. These may have played an important role for some of the teams, but they could not be monitored and analyzed. However, students were informed that their contributions on blog and Twitter would be used to follow and evaluate their collaboration throughout the semester, and as a way of documenting the progress of the project. Hence, a large part of student communication did take place on these social media tools. We therefore argue that the blog and microblog posts provide a reliable reflection of the community of inquiry formed by the students.

The study sample, while not very large (75 students who generated 479 blog entries and 1233 tweets), is more substantial compared to similar reports. Thus, there are 21 students and 131 blog posts and comments analyzed in (Angelaina & Jimoyiannis, 2012a), 15 students and 87 blog comments in (Pifarre et al., 2014) and 47 students and 324 tweets in (Sinnappan & Zutshi, 2011). Furthermore, all posts were analyzed by two coders, with high inter-reliability rates; CollAnnotator support tool was used, instead of a manual approach for content analysis. This provided essential features such as: support for multiple coders and the negotiation process, comprehensive annotation functionality, support for multiple categories per unit of analysis, detailed statistics and reports with graphical visualizations, all in an intuitive and easy to use interface.

As future work, we plan to investigate also the contribution of individual students / teams to the community of inquiry, based on the functionalities offered by CollAnnotator. Indeed, the tool provides support for more in-depth analyses, by generating reports at student and team level. All the tables and charts computed for the whole community are also generated for each individual learner and each team. We could thus investigate the profiles of individual students / teams, the proportion of each presence they exhibit and their contribution for the construction and maintenance of the community of inquiry. Furthermore, exploring the relationship between the presences exhibited by the students and their learning outcomes would be a valuable endeavor.

The investigation could be extended also with social network analysis (as in Jimoyiannis and Angelaina (2012), Shea et al. (2013), and Shea et al. (2014)) or learning mapping (as in Angelaina and Jimoyiannis (2012b)). We have already designed a knowledge extraction framework for a social learning environment and used SNA to investigate students' collaboration patterns in eMUSE platform, for a different student cohort (Becheru et al., 2018). Furthermore, alternative coding schemes could be integrated in CollAnnotator, to provide a more comprehensive analysis of the student generated content (e.g., discourse categories proposed in Fu et al. (2016), and Ioannou et al. (2015)). Finally, extending the study to different instructional scenarios and social media settings in various educational contexts would be a worthwhile research direction.

References

Anderson, T. (2019). Challenges and opportunities for use of social media in higher education. *Journal of Learning for Development*, *6*(1), 6-19.

Angelaina, S., & Jimoyiannis, A. (2012a). Educational blogging: Developing and investigating a students' community of inquiry. In A. Jimoyiannis (Eds.), *Research on e-Learning and ICT in education* (pp. 169-182). New York, NY: Springer.

Angelaina, S., & Jimoyiannis, A. (2012b). Analysing students' engagement and learning presence in an educational blog community. *Educational Media International, 49*(3)*,* 183-200.

Arbaugh, J. B., Cleveland-Innes, M., Diaz, S. R., Garrison, D. R., Ice, P., Richardson, J. C., & Swan, K. P. (2008). Developing a Community of Inquiry instrument: Testing a measure of the Community of Inquiry framework using a multi-institutional sample. *Internet and Higher Education, 11*(3-4), 133-136.

Ardaiz-Villanueva, O., Nicuesa-Chacon, X., Brene-Artazcoz, O., Sanz de Acedo Lizarraga, M. L., & Sanz de Acedo Baquedano, M. T. (2011). Evaluation of computer tools for idea generation and team formation in project-based learning. *Computers & Education, 56*(3), 700-711.

Badea, G., & Popescu, E. (2017). CollAnnotator - A Support tool for content analysis according to Community of Inquiry framework. In M. Chang et al. (Eds.), *Proceedings IEEE 17th International Conference on Advanced Learning Technologies - ICALT 2017* (pp. 212-214). doi:10.1109/ICALT.2017.127

Becheru, A., Calota, A., & Popescu, E. (2018). Analyzing students' collaboration patterns in a social learning environment using StudentViz platform. *Smart Learning Environments, 5*, Article 18. doi:10.1186/s40561-018-0063-0

Burgess, M. L., Slate, J. R., Rojas-LeBouef, A., & LaPrairie, K. (2010). Teaching and learning in second life: Using the community of inquiry (CoI) model to support online instruction with graduate students in instructional technology. *Internet and Higher Education, 13*(1-2), 84-88.

Creemers, B., & Kyriakides, L. (2010). School factors explaining achievement on cognitive and affective outcomes: Establishing a dynamic model of educational effectiveness. *Scandinavian Journal of Educational Research, 54*(3), 263-294.

Eteokleous, N., Ktoridou, D., & Orphanou, M. (2014). Integrating wikis as educational tools for the development of a Community of Inquiry. *American Journal of Distance Education, 28*(2), 103-116.

Farrow, E., Moore, J., & Gašević, D. (2019). Analysing discussion forum data: A Replication study avoiding data contamination. In R. Ferguson et al. (Eds.), *Proceedings of the 9th International Conference on Learning Analytics & Knowledge - LAK19* (pp. 170-179). New York, NY: ACM.

Ferreira, M., Rolim, V., Ferreira Mello, R., Dueire Lins, R., Chen, G., & Gašević, D. (2020). Towards automatic content analysis of social presence in transcripts of online discussions. In V. Kovanovic et al. (Eds.), *Proceedings of the 10th International Conference on Learning Analytics & Knowledge - LAK20* (pp. 141-150). New York, NY: ACM.

Fu, E., van Aalst, J., & Chan, C. (2016). Toward a classification of discourse patterns in asynchronous online discussions. *International Journal of Computer-Supported Collaborative Learning, 11*, 441-478.

Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education, 2*(2-3), 87-105.

Garrison, D. R., Anderson, T., & Archer, W. (2010). The First decade of the Community of Inquiry framework: A Retrospective. *Internet and Higher Education, 13*(1-2), 5-9.

Garrison, D. R., Cleveland-Innes, M., Koole, M., & Kappelman, J. (2006). Revisiting methodological issues in transcript analysis: Negotiated coding and reliability. *Internet and Higher Education, 9*(1), 1-8.

Hayes, S. (2014). *A Mixed methods study of shared epistemic agency in team projects in an online baccalaureate nursing course.* [Doctoral thesis, State University of New York at Albany]. PQDT Open. https://pqdtopen.proquest.com/pubnum/3667396.html

Hayes, S., Uzuner-Smith, S., & Shea, P. (2015). Expanding learning presence to account for the direction of regulative intent: self-, co- and shared regulation in online learning. *Online Learning, 19*(3), 15-31.

Ioannou, A., Brown, S., & Artino, A. (2015). Wikis and forums for collaborative problem-based activity: A Systematic comparison of learners' interactions. *Internet and Higher Education, 24*, 35-45.

Jimoyiannis, A., & Angelaina, S. (2012). Towards an analysis framework for investigating students' engagement and learning in educational blogs. *Journal of Computer Assisted Learning, 28*(3), 222-234.

Jimoyiannis, A., & Tsiotakis, P. (2017). Beyond students' perceptions: investigating learning presence in an educational blogging community. *Journal of Applied Research in Higher Education, 9*(1), 129-146.

Kazanidis, I., Pellas, N., Fotaris, P., & Tsinakos, A. (2018). Facebook and Moodle integration into instructional media design courses: A comparative analysis of students' learning experiences using the Community of Inquiry (CoI) model. *International Journal of Human–Computer Interaction, 34*(10), 932-942.

Kim, P., Hong, J. S., Bonk, C., & Lim, G. (2011). Effects of group reflection variations in project-based learning integrated in a Web 2.0 learning space. *Interactive Learning Environments, 19*(4), 333-349.

Kreijns, K, Van Acker, F., Vermeulen, M., & Van Buuren, H. (2014). Community of Inquiry: Social presence revisited. *E-Learning and Digital Media, 11*(1), 5-18.

Kuo, Y.-C., Belland, B. R., & Kuo, Y.-T. (2017). Learning through blogging: students' perspectives in collaborative blog-enhanced learning communities. *Educational Technology & Society, 20*(2), 37-50.

Lumby, C., Anderson, N., & Hugman, S. (2014). Apres le deluge: Social media in learning and teaching. *The Journal of International Communication, 20*(2), 119-132.

McKerlich, R., & Anderson, T. (2008). Community of inquiry and learning in immersive environments. *Journal of Asynchronous Learning Networks, 11*, 35-52.

Neto, V., Rolim, V., Ferreira, R., Kovanović, V., Gašević, D., Dueire Lins, R., & Lins, R. (2018). Automated analysis of cognitive presence in online discussions written in Portuguese. In V. Pammer-Schindler et al. (Eds.), *Lifelong Technology-Enhanced Learning. EC-TEL 2018* (pp. 245-261). doi:10.1007/978-3-319-98572-5_19

Öztürk, E. (2015). Facebook as a new community of inquiry environment: an investigation in terms of academic achievement and motivation. *Journal of Baltic Science Education, 14*(1), 20-33.

Pifarré, M., Guijosa, A., & Argelagós, E. (2014). Using a blog to create and support a community of inquiry in secondary education. *E–Learning and Digital Media, 11*(1), 72-87.

Popescu, E. (2014). Providing collaborative learning support with social media in an integrated environment. *World Wide Web*, *17*(2), 199-212.

Popescu, E., & Badea, G. (2017). Using CollAnnotator to analyze a Community of Inquiry supported by educational blogs - Preliminary results. In É. Lavoué et al. (Eds.), *Data Driven Approaches in Digital Education. EC-TEL 2017* (pp. 580-583). doi:10.1007/978-3-319-66610-5_67

Popescu, E., & Petrosanu, L. M. (2016). Integrating a peer evaluation module in a social learning platform. In E. Popescu et al. (Eds.), *Innovations in Smart Learning* (pp. 141-150). Singapore: Springer.

Remesal, A., & Friesen, N. (2014). Inquiry into 'Communities of Inquiry': Knowledge, communication, presence, community. *E–Learning and Digital Media, 11*(1), 1-4.

Ricoy. M. C., & Feliz, T. (2016). Twitter as a learning community in higher education. *Educational Technology & Society, 19*(1), 237-248.

Savery, J. R., & Duffy, T. M. (1995). Problem-based learning: An Instructional model and its constructivist framework. *Educational Technology, 35*(5), 31-38.

Shea, P., & Bidjerano, T. (2010). Learning presence: Towards a theory of self-efficacy, self-regulation, and the development of a communities of inquiry in online and blended learning environments. *Computers & Education, 55*(4), 1721-1731.

Shea, P., & Bidjerano, T. (2012). Learning presence as a moderator in the community of inquiry model. *Computers & Education, 59*(2), 316-326.

Shea, P., Hayes, S., Uzuner-Smith, S., Gozza-Cohen, M., Vickers, J., & Bidjerano, T. (2014). Reconceptualizing the community of inquiry framework: An Exploratory analysis. *Internet and Higher Education*, *23*, 9-17.

Shea, P., Hayes, S., Uzuner-Smith, S., Vickers, J., Bidjerano, T., Gozza-Cohen, M., Jian, S. B., Pickett, A., Wilde, J., & Tseng, C. H. (2013). Online learner self-regulation: learning presence viewed through quantitative content- and social network analysis. *International Review of Research in Open and Distance Education*, *14*(3), 428-461.

Shea, P., Hayes, S., Uzuner-Smith, S., Vickers, J., Bidjerano, T., Pickett, A., Gozza-Cohen, M., Wilde, J., & Jian, S. (2012). Learning presence: Additional research on a new conceptual element within the community of inquiry (CoI) framework. *Internet and Higher Education, 15*(2), 89-95.

Shea, P., Hayes, S., Vickers, J., Gozza-Cohen, M., Uzuner, S., Mehta, R. et al. (2010). A Re-examination of the community of inquiry framework: Social network and content analysis. *Internet and Higher Education, 13*(1-2), 10-21.

Sinnappan, S., & Zutshi, S. (2011). Using microblogging to facilitate community of inquiry: An Australian tertiary experience. In G. Williams et al. (Eds.), *Changing Demands, Changing Directions. Proceedings ASCILITE 2011* (pp. 1123-1135). Hobart, Australia: Australasian Society for Computers in Learning in Tertiary Education.

Swan, K. (2019). Social construction of knowledge and the Community of Inquiry framework. In I. Jung (Eds.), *Open and Distance Education Theory Revisited* (pp. 57-65). Singapoe: Springer.

Swan, K., & Ice, P. (2010). The Community of inquiry framework ten years later: Introduction to the special issue. *Internet and Higher Education, 13*(1-2), 1-4.

Traver, A., Volchok, E., Bidjerano, T., & Shea, P. (2014). Correlating community college students' perceptions of community of inquiry presences with their completion of blended courses. *Internet and Higher Education, 20*, 1-9.

Wertz, R. (2014). *Toward a new model within the community of inquiry framework: Multivariate linear regression analyses based on graduate student perceptions of learning online* (Unpublished doctoral dissertation). Purdue University, West Lafayette, ID.

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

Xing, W., & Gao, F. (2018). Exploring the relationship between online discourse and commitment in Twitter professional learning communities. *Computers & Education, 126*, 388-398.

Yang, J. C., Quadir, B., Chen, N. S., & Miao, Q. (2016). Effects of online presence on learning performance in a blog-based online course. *Internet and Higher Education, 30*, 11-20.

Yeh, H.-C., Tseng, S.-S., & Chen, Y.-S. (2019). Using online peer feedback through blogs to promote speaking performance. *Educational Technology & Society, 22*(1), 1-14.

Zacks, S., & Hen, M. (2018). Academic interventions for academic procrastination: A Review of the literature. *Journal of Prevention & Intervention in the Community, 46*(2), 117-130.

Chen, S., Zhang, S., Qi, G. Y., & Yang, J. (2020). Games Literacy for Teacher Education: Towards the Implementation of Game-based Learning. *Educational Technology & Society*, *23* (2), 77–92.

Games Literacy for Teacher Education: Towards the Implementation of Game-based Learning

Si Chen[1], Sujing Zhang[1], Grace Yue Qi[2*] and Junfeng Yang[1]
[1]Hangzhou Normal University, China // [2]Massey University, New Zealand // sylviachen327@gmail.com // 21sjzhang@163.com // G.Qi@massey.ac.nz // yangjunfengphd@gmail.com
[*]Corresponding author

(Submitted March 25, 2020; Revised June 27, 2020; Accepted July 8, 2020)

ABSTRACT: Game-based learning (GBL) has been widely recognised in research, and evidently benefited for learners. However, what GBL is perceived by teachers and learners has been a concern that might impact on quality of teaching and learning in the GBL environment. Game-based pedagogy meticulously designed from a teacher's perspective was regarded as harping on the same string without fun by learners. This paper aims to explore games literacy capabilities in supporting teachers to implement GBL that meets learners' needs and expectations. Semi-structured interviews and surveys with experienced teachers of GBL and experts in the relevant field were conducted, followed by an Analytic Hierarchy Process seeking perceptions of a group of academics and researchers. Findings suggested five key capabilities in game literacy required by teachers in implementing GBL. They are (1) basic games literacy, (2) high-level games literacy, (3) instructional design for GBL, (4) organisation and management for GBL, and (S) evaluation of GBL. Amongst the five, instructional design for GBL and high-level games literacy were rated highly impacting on the quality of teaching. Based on the findings, aiming at informing teacher education and professional development, we proposed a framework providing a guidance to improve game-based design and pedagogical practices for teachers in the implementation of GBL in their classrooms. It concludes that teachers' capabilities in games literacy require specific attention to instructional design – that demands a thought-provoking process for GBL.

Keywords: Game-based learning, Games literacy, GLTE framework, Instructional design, Teacher education

1. Introduction

Over the past decades, there has been a growing attention for the use of digital games in learning and instructions, often referred to as serious games or game-based learning (GBL). In this respect, two dimensions of learning are regarded to contribute to GBL: cognitive and affective (O'Neil, Wainess, & Baker, 2005). Both dimensions of learning emphasise the adaptation of learning to accommodate learners' cognitive needs and interest, and provide motivation for learning (Malone, 1981).

Previous studies have shown that GBL is more effective than conventional instructions, such as lectures or classroom instruction (e.g., Sitzmann, 2011; Wouters & van Oostendorp, 2013). However, it is argued that GBL often involves complex learning environments, where players can be easily overwhelmed by information and activities provided for learning (Wouters & van Oostendorp, 2013). This raises a question that whether GBL environments can afford learners to engage in processing cognitive activities as it requires a careful instructional design and support (e.g., scaffolding, giving feedback etc.) from teachers as designers and facilitators (Wouters & van Oostendorp, 2013).

There is also a debate that GBL is perceived by teachers as an effective approach in educational practices. Teachers often acknowledge the merits of games whereas they complain the difficulty of completing effective instructional design, which seriously hinders the advantages of GBL. Merrin (2009) believed that new media posed a threat to teachers' authority in practice. One of the reasons may relate to differentiated interests and understandings of GBL. Although many teachers are confident in developing computing skills as presented in the "Serious Play" project (Beavis, 2017), a fear over adopting digital technologies to teaching does exist (Pivec, 2009; Zhu & Wang, 2018).

Furthermore, it is believed that many teachers separate "digital technologies" from GBL as they tend not to see it as an entity for learning and teaching (del Blanco et al., 2012; Yukselturk, Altiok & Başer, 2018). This indicates constructivist pedagogical approaches which often emphasise games enabled authentic learning opportunities where learners are agents in the process (Mama & Hennessy, 2013). It is most likely that teachers use the same pedagogical practices for the use of provided tools or platforms like other tools or resources in their classroom (Prestridge, 2017).

A systematically conceptualised games literacy is much needed to equip teachers in achieving GBL. This is particularly discussed in Zimmerman's (2009) work, where he argued that games design is a paradigm for understanding and addressing the key components of digital literacy. It is essential that a theory-based practical measurement should be developed to instruct literacy diagnosis and educational strategies in the implementation of GBL (Klimmt, 2009). According to Zimmerman's (2009) three concepts of gaming literacy (which we prefer "games literacy" in this paper), namely systems, play, and design, teachers' literacy on games and capability development of understanding the dynamics of games are key to achieve GBL.

Given all the concerns, the present study aims to explore what GBL informed games literacy for teachers is and what teachers think about their needed capabilities in games literacy for the implementation of GBL in the classroom. It is notable that teachers' capability development in games literacy caters for learner's needs and expectations in a GBL environment (Osatuyi et al., 2018). In line with the principle and value of the digital game culture, the role of teachers needs to be explicitly expressed when GBL is usually co-designed and conducted with learners. We aim to address the following research questions by conducting this study:

- Research question 1: What are the essential capabilities in games literacy for teachers informed by GBL?
- Research question 2: What are the most important capabilities in games literacy that teachers need to prioritise in the implementation of GBL?

The exploration of the key capabilities in games literacy informed by GBL for teachers will enable us to propose a framework providing a guidance to improve game-based design and pedagogical practices informing teacher education and professional development in the GBL pedagogy. As discussed briefly, there are many terms and concepts relative to GBL hence how we define and use them in this study is very crucial.

2. Literature review

2.1. Defining terms: GBL and games literacy

The term GBL mostly emphasises a type of gameplays for learners during learning and achieving positive outcomes (Shaffer, Halverson, Squire & Gee, 2005). In other words, this term identifies learners' ownership and student-centred learning experience. In a technological approach, game-based teaching is often used as differing from the traditional teaching approach (Jackson, 2009). This relies on instructors' design in terms of the nature of games and learning objectives in line with curricula (Becker, 2017). Therefore, in this paper we specifically focus on GBL as it highlights a student-centred approach.

According to Day (1973), the practice and understanding of games literacy was limited to the observation of games. It was believed that games literacy was an ability that players must build through multi-plays of the games in a successively changing environment, including game scenes and various platforms. In other words, games literacy was highlighted for its functionality and usefulness contributing to sound gaming experience for learners (Day, 1973). Subsequently, games literacy was regarded as a distinctive literacy similar to the television, film and literature, belonging to cultural media literacy (Buckingham & Burn, 2007). In that, games literacy, instead of being understood as a theory, was more appropriately derived from three dimensions, namely culture, criticism, and creativity (Partington, 2010). Although the content of games literacy varies in contexts it is emphasised in different aspects (Kim, 2008; Wang, 2010; Zagalo, 2010).

Meanwhile, the functions of games literacy also vary in addressing challenges of the recent developments in the digital games industry (Klimmt, 2009). To be specific, games literacy as an effective mechanism could potentially protect (adolescent) players from undesirable gaming effects, avoiding the impact of gaming violence (Cantor & Wilson, 2003) or preventing youths from games addiction (Tao, 2009). In the field of education, based on Gee's notion of literacy (Gee, 2003), Zagal and Jose (2008) developed a framework implemented as an assessment tool of developing students' games literacy competence.

Apperley and Beavis (2013) developed a model for critical games literacy in which they defined the gaming literacy instead of games literacy as games are enacted by players, mostly students in educational contexts. As new literacies, they are associated with traditional 'literacy' such as reading and writing. This implies that gaming literacy involves "texts" as well as "action" which is believed that gaming literacies (Zimmerman, 2009) were informed in actions of digital gameplays (Atkins, 1993). Zimmerman (2009) specified three concepts of gaming/games literacy: systems, play and design, which highlighted the role of teachers in the process. However little research has examined the importance of teachers' games literacy and essential capabilities they need for designing, implementing, and facilitating game-based learning in various educational settings.

2.2. Games literacy for learners

Nowadays, 21st century skills have been stressed, which urges the change to a third perspective to observe games literacy, that is games literacy could be of social and cultural capitals much needed for young learners (Bourgonjon, 2014; Partington, 2010; Yeh, Chang & Chen, 2019). In addition, it is seemingly that it develops a necessary systematic, critical thinking skill demanded for learners (Salen, 2007) and other new sets of ideas and skills as well as practices (Zimmerman, 2009).

Some scholars assert that games literacy is necessary to be acquired by players/learners, and received recognition or understanding of by people around, including parents and friends (e.g., Robertson, 2018). Chuang, Chen, Chen, Shen and Tsai (2011) pointed out that the understanding of games literacy and its emerging advantages for learners should be explained to parents in order to avoid potential family cognitive conflicts regarding digital gameplays. It is important to address scaffolding that is typically provided by parents and teachers to engage with children/students for effective interactions for learning purposes in informal or formal context respectively (Tollånes, Aarsand & Sandberg, 2011).

It is notable that not all learners are games players, and not all teachers are keen advocates for the use of games or experienced with adopting games to engage students with learning in the classroom (Beavis, 2017). Therefore, it is important to understand what capabilities are required for teachers to at least attempt to experience the GBL pedagogy.

2.3. Games literacy for teachers

Kali, McKenney and Sagy (2015) suggested that teachers should be seen as designers of digital learning with the aid of technology. McKenney (2005) synthesised "ecological" as the term to describe the complex and dynamic settings in which teachers have to work in designing learning for learners. In addition to "design", teachers are also facilitators of learning when it occurs (Carey, 1993). Given the strength of the learners have in terms of games literacy, what teachers should do to design and facilitate such GBL pedagogical learning is vital (Hsu, Tsai, Chang & Liang, 2017).

Catherine Beavis and her colleagues (Beavis, 2017) developed an Australian Research Council funded study on literacy teachers' beliefs in "Serious Play" (2015-2017). They adopted a term "Serious Play" comprising of educational games or game-based activities for both informal and formal learning. Rowan and Beavis (2017) reported that those literacy teachers interviewed expressed a clear belief in the potentially successful educational outcomes. In light of their belief, they showed confidence of innovation within an existing curriculum and inside the boundaries provided by different education authorities, schools and individuals. Researchers indicated teachers' pedagogical beliefs typically expressed in three areas: developing computer skills, delivering the content of the curriculum, and enacting change in teaching and learning (Prestridge, 2017). These imply that teachers engaging with digital games rely on their belief and roles in student learning (Prestridge & de Aldama, 2016).

Drawing on two case studies, Beavis and O'Mara (2010) believed that successful gameplay entails simultaneous attention to a number of elements in the use of a wide range of literacy practices which go well beyond what most literacy required in subject-based curricula. Crucially, Beavis (2013) emphasises active and creative dimensions of work with texts from a design perspective. She warned that teachers should look forward not backward when reimagining curricula to engage with texts and literacies of "old times." Teachers should consider that games play an important role as challenging but important hybrid textual forms that are closely linked with action.

There is a scarcity of research looking into teachers' games literacy but in fact, it can impact on the uptake and implementation of GBL and discouraging young learners' interest in and engagement with GBL. According to the review above, our proposed framework refers to the knowledge and emotion of teachers in the process of GBL, and the competence to strike a balance between games, teaching, and students by endorsing game thinking. To provide clear guidelines to teachers in supporting their games literacy capabilities development, this study investigates the essential and prioritised factors drawn attention to experienced teachers. We take these into account to develop an emerging framework on games literacy for teacher education and professional development as required in the implementation of GBL.

3. Methodology

3.1. Research design

The present research aims to explore what GBL informed games literacy for teachers is and what teachers think about their needed capabilities in games literacy for the implementation of GBL in the classroom. The design of the present research is essential, where teachers were the key participants to study, including their perceptions and experiences of GBL. Employing a combination of quantitative and qualitative research (Creswell, 2012), the study contains three phases of research in which semi-structured interviews (Phase One), a quantitative systematic review of literature (Phase One), one online survey (Phase Two) and one email survey (Phase Three) were conducted.

A critical evaluation of the findings was followed and led to proposing an emerging framework for teachers to develop games literacy in the GBL environment. The study sought to answer the two research questions (see Section 1). Both questions are answered in the findings of the study and Research Question 2 is further discussed in constructing the proposed framework.

3.2. Data collection

This study involved three phases of data collection: (1) a quantitative systematic review of literature, and interviews with teachers, (2) online survey with selected teachers and researchers, and (3) email survey with experts and scholars in the field of educational technology and teacher education.

3.2.1. Phase One: Semi-structured interviews and a quantitative systematic review of literature

Interview participants: The first phase of data collection, semi-structured interviews, recruited two groups of teachers and researchers. One was out reached in March 2018 where eight participants who were awardees of Competition in Educational Games 2016-2017 were invited to join one-on-one tele-interviews. The other one involved six participants including five teachers and one principal from various primary and secondary schools who were contacted in May 2018. These participants were the members of a distinguished teacher group and they often delivered professional development workshops addressing GBL to other teachers.

Design of the interviews: A selective sampling was adopted when choosing the interview sample as the objects that can provide insights in relation to research questions (Bernard, 2017). It was aimed at collecting and completing key items of games literacy which were evidently required in the implementation of GBL from a practical perspective. Besides bio-data related interview questions, two principle topics as follows were designed in the protocol. All interviews were recorded with a total of 240 minutes. No demographic information was aimed to gather as the focus of the interviews was only the questions presented below:

- If you were invited to give a lecture for teachers about GBL, what procedures would you follow?
- What capabilities are required for teachers to design and deliver GBL?

A systematic review of indexed journal articles: the research team conducted a systematic review in a quantitative approach. The aim is to examine any missing relevance of “Games Literacy” appeared in the indexed SSCI and AHCI journals. This step complemented the results of the interviews and assisted in defining key concepts of games literacy.

3.2.2. Phase Two: Online survey

Participants: An Online Survey was designed and distributed to experienced GBL teachers and researchers who attended the Game-based Teaching Excellence Class Exhibition – the 3rd Game-based Learning International Conference. The research team generated an Online Survey QR code and invited all conference participants to fill out the survey. The team also held a couple of presentations where they were able to embed the QR code in the presentation slides to recruit participants. The team also approached teachers who they had known and attended the conference to complete the online survey. This event provided the research team with a great opportunity to reach out experienced teachers and researchers of GBL.

Design: In accordance with the reviewed literature and results from the interviews, we compiled the items collected in the first phase into a survey. The purpose is to use the results of exploratory factor analysis (EFA) (Yong & Pearce, 2013) to classify and reduce the dimensionality of all games literacy to make it more logical. The Phase Two survey was an online survey accessible from July 2018 to January 2019. The survey received a total of 240 respondents. After eliminating the invalid responses, 231 were included for analysis. Invalid responses were defined as those incomplete and/or with a consistent single selection throughout. The effective survey recovery rate was 95%. The Cronbach's alpha was $\alpha = 0.948$.

Every question in this survey was designed in a five-point Likert scale in the form of "1 = strongly disagree" to "5 = strongly agree." Questions were all delivered in the formatted expression starting with "I think teachers who adopt game-based teaching should be of…". There were 39 items presented in the survey including sections such as, a brief introduction, a description of game-based learning, 5 bio-data questions, and 34 research variables seeking perceptions of respondents (see Appendix B).

To examine the validity of the survey, instructors from a primary school in Futian District, Shenzhen, China helped respond to pilot survey questions. The instructors were expected to provide feedback on the feasibility and sensibility of each question. Based on the feedback from a total of 50 samples, the original version of the survey was modified. Of five questions were removed from further analysis as the score of Corrected Item-Total Correlation (CITC) was tested smaller than 0.6. Whereas the retained items were refined based on the opinions.

3.2.3. Phase Three: Email survey

Participants: A relevant comparison matrix in accordance with Analytical Hierarchical Process (AHP) was designed (see Appendix D). In that the second survey was aimed to seek knowledge and perceptions of experts and scholars on games literacy. These experts worked in the field of educational technology and teacher education. The Phase Three survey was distributed through emails. In total, 20 experts and scholars from highly reputable higher education institutions in China were contacted to provide consolidated opinions. The response rate was 100%.

Design: The main purpose of this phase was to consolidate professional opinions of experts in GBL and identify the most important games literacy factors and capabilities needed to be prioritised for teachers to develop given their busy schedule. AHP was used here to make a pairwise comparison of literacy elements, and to judge their relative importance in the form of scores. Specifically, it helped identify the key literacy that instructors can prioritize to improve teaching quality, which is beneficial to evaluation and self-examination. Merits on the importance of these factors can be done by giving weights to each one. The results of the survey contribute responses to the RQ2 identifying the most important capabilities that should be prioritised by teachers to develop in games literacy.

3.3. Data analysis

Each phase was undertaken differently, therefore analytical tools and approaches varied due to the nature of their designs aligned with the research questions.

3.3.1. A systematic review and interview data analysis

The "Sougou Hearing" voice transfer software was used to transcribe the interview recordings. Four research assistants manually checked recordings to ensure the alignment of the voice text and the voice content. These research assistants were coded as A, B, C and D and worked systematically; for instance, A was responsible for extracting key words from the transcribed texts, which followed by B and C's reviews and D's finalisation.

An extensive literature review was conducted to search for items that might not be mentioned in the interview, and to ensure the factors of GBL completely converted. Researchers used Games Literacy as the keyword to search for peer-reviewed scholarly articles listed in SSCI and AHCI indexes. Eighteen refereed articles and two books containing a clear claim about items of games literacy were incorporated for analysis. 19 key words were identified and added to our list of correlated items generated in the first online survey (see Appendix B).

3.3.2. Online survey data analysis

The online survey data were loaded to SPSS version 24.0 and analysed starting from an internal consistency of the instrument. The data were entered into the SPSS software to produce descriptive statistics focusing on frequency, means, standard deviations and percentages. Exploratory Factor Analysis (EFA) was employed to generate factors (or components) demonstrating the relationships among the statement questions (variables). A couple of initial tests, such as descriptive statistics and KMO and Bartlett's Test, were conducted to check if it was worthy of factor analysis. Then, a varimax method of factor rotation was used along with multiple loadings to ultimately generate main factors underlying statistical validity. Each factor interpreted grouped items (or variables) for a clear report of the data.

To ensure the reliability of the generated factors, Cronbach's alpha (α) was used to indicate the overall reliability of the scale. It was then followed by an EFA that was conducted to identify the underlying correlation within the 39 items which are also key factors to GBL. Each factor's reliability testified with Cronbach's alpha was examined. An alpha level of 0.05 was used for the statistical tests.

3.3.3. Analytic Hierarchy Process (AHP) for email survey

Analytic Hierarchy Process (AHP) is a multi-criteria decision-making method combining qualitative and quantitative analysis. It allows to take simple pairwise comparison judgments of experts to determine overall priorities in order to rank the alternatives (Saaty, 2008). AHP has been used in various settings to make decisions and such analytical method has been adopted and written by scholars (e.g., Bhushan & Ria, 2004; Hummel, 2001).

To complete a hierarchical analysis, a structure of the problem as a hierarchy was first established which was acquired from the result of the EFA that was employed to analyse the findings of the Phase One online survey. The identified five factors were then divided from two dimensions as the results of the factor analysis showed a distinctive property between generating games literacy (basic games literacy, high-level games literacy) and practical game-based learning literacy (instructional design of GBL, organisation and management of GBL, evaluation of GBL).

Based on the findings of the previous two phases, a three-level hierarchy containing 2 dimensions, 5 constructs and 18 items were emerging as criteria to examine the importance of the factors in a given level with respect to some or all of the items in the adjacent level above. The study then followed the 1~9 scale method specified in the AHP. A three-level hierarchy as the form of the matrix was established, which was followed with an elicit pairwise comparison assessment (see Figure 1).

In terms of data processing, we used a software that specializes in processing AHP data. This software was developed by a technology company, and we purchased the license to use it. Therefore, the data processing was assisted by computer software according to the principle of AHP.

Figure 1 illustrates the calculation idea. The last step was to use the formula to sort out the factors in the matrix. The Formula 3-1 (see below) was used to calculate the scoring results, and the relative importance weights between the lowest layer and the highest layer in turn. Formula 3-2 was used to calculate consistency ratio (CR.) – test the consistency of the matrix on individual perceptions. Random index (RI.) can be obtained by referring fixed figures (see Appendix C).

$$\omega_i = \frac{1}{n}\sum_{j=1}^{n}\frac{a_{ij}}{\sum_{k=1}^{n}a_{kj}} \quad \text{Formula 3-1}$$

$$\mathrm{CR.} = \frac{CI.}{RI.} \quad \text{Formula 3-2}$$

Figure 1. Formula of Analytic Hierarchy Process (AHP)

4. Results

4.1. Results from interviews and a systematic review

As explained in the analytical process, both the interviews with two groups of teachers and the keyword "Games Literacy" search in the scholarly articles indexed in SSCI and AHCI contributed to the correlated items in relation to essential capabilities in games literacy for teachers.

19 capabilities were mentioned in the interviews and literature search which were emerging into 6 sections consisting of 39 survey questions:

game knowledge, game belief, game spirits, game thinking, game awareness, game morality, game critique, game creativity, game access, game operation, game identity, activity design of GBL, curriculum design of GBL, product design of GBL, organisation and implementation of GBL, the gamification management of classes and schools, assessment of game-based activities, assessment of game-based curriculum, assessment of game-based environment.

4.2. Demographic information in Phase Two's online survey

Demographic characteristics of the respondents were examined. Among the 231 respondents, 64.0% were male and 36.0% were female. In terms of years of teaching experience, 2.7 years was reported as a medium. Instructors in Information Technology disciplines accounted for a maximum of 28.4%, followed by instructors in Others (23.46%) who were specialised in subjects like Educational Technology and Mobile Learning – are the common subjects in the universities of China.

About 32.2% indicated that they were employed in higher education, followed by instructors who were engaged with teaching projects in primary schools (22.22%) and high schools (22.22%). It could be understood that GBL is widely focused in tertiary level since tertiary students are often mature in handling the relationship between technique-related games and formal learning. In addition, approximately 23.46% reported that they were also teaching other disciplines.

Moreover, GBL is more commonly accepted by younger-aged teachers that is related to their stress level at work and willingness of adopting digital technologies. More than half of participants (65%) reported that they did not play games every week, alternatively spending less than 2 hours on games weekly.

4.3. Results from Exploratory Factors Analysis (EFA) on Phase Two's online survey

The EFA extraction method used on the scales was Principal Axis Factoring. The rotation method selected was Varimax. The level of commonality among the items was considered wide (from 0.6-0.8) scales. The Principal Component Analysis was used to help establish the number of factors to retain. Factors are often retained if the eigenvalue for the actual data is larger than 1.0 (Kaiser, 1974). All the factors retained for scales had an eigenvalue > 1. It was determined through these criteria that the five factors indicated by the Principal Component Analysis were retained, accounting for 71.550% of the variance (Table 1).

Table 1. Adoption factors total variance explained and reliability

Factors	Eigenvalues			Number of items	Cronbach's alpha
	Total	% of variance	Cumulative variance		
1	10.967	21.612	47.684	7	0.901
2	1.984	16.307	56.311	6	0.878
3	1.282	14.590	61.885	4	0.865
4	1.172	11.337	66.982	3	0.854
5	1.051	7.309	71.550	3	0.844
6	0.787	3.422	74.973		

With the 5 factors, the Kaiser-Meyer-Olkin measure of sampling adequacy was 0.878, which is above the minimum threshold of 0.5 suggested by Kaiser (1974). Bartlett's test of sphericity had a Chi square of 4187.751 ($p < .001$), indicating that the intercorrelation matrix contained variables with sufficient collinearity for analysis (Bartlett, 1950). The rotation converged in 9 iterations. The factor loadings were evaluated through using the items and the weight of their loadings on each factor. A descriptive label was assigned to each factor. Table 1 contains the variance in relation to each factor, the enclosed number of items and each factor's Cronbach's alpha.

The first factor contained 7 items from the 5 concepts in GBL, such as activity design, curriculum design, product design, game creation and design of such environment. These items are related to the procedures of design when teachers integrate games (game-related elements) into classroom, including the design of environment, the design of learning activities and the design of assessments. The term Instructional Design (ID) refers to instructional system development (ISD), namely the stages of analysis, design, development, implementation and evaluation (Dick, Carey & Carey, 2005). It involves the utilization of strategies, assessments and techniques to integrate educational resources to accomplish an instructional design. As the evaluation of learning needs to be gamified, some game-related elements, regarded as normal assessment tools, such as scoreboard, ranking and reward need to be embedded before introducing them while teaching in the classroom. Therefore, this factor was named as Instructional Design of GBL accounting for 21.612% of the variance, indicated that this is a very important construct of the emerging GLTE framework (details can be found in Section 5).

The second factor contained 6 items from the 4 concepts of game belief, game critique, evaluating GBL, and evaluating learning with gamification. These items related to issues of distinctive games, keeping a critical attitude towards games, and evaluating GBL. Therefore, this factor was named as evaluation of GBL. There are two different connotations to this literacy construct: one refers to an evaluation method involving game-related elements in the evaluation; and the other refers to an object which can be evaluated by educators (Chen & Zhang, 2019).

The third factor contained 4 items from the 4 concepts of game knowledge, game manipulation, game morality and game access. These items relating to issues of the basic qualities are necessary for better gameplays, implying that without this literacy element players are hardly exposed to a good gaming experience. Therefore, this factor was named as basic games literacy.

The fourth factor included 4 items from 3 concepts of game thinking, game identity and game spirits, involving the application of games and resolution to real life problems. Games literacy takes the developmental processes and social personal resources into account. The social personal resources can be explained as cognitive skills and social support from friends which is contrary to states of psychological crisis (Klimmt, 2009). Therefore, this factor was named high-level games literacy.

The fifth factor is relevant to the following three contexts: applying designed game-based instructions into practice, organizing GBL in the classroom, and using gamification to manage learning and teaching. This factor contained three items from the two conceptual domains: organisation and implementation of GBL, and the gamification management of classes and schools. Therefore, this factor was named as organisation and management of GBL.

The first research question explores what literacy factors are necessarily equipped by teachers. This was mainly responded by conducting the EFA involved in the Phase One online survey. The results demonstrated the following five game literacies: basic games literacy, high-level games literacy, instructional design of GBL, organisation and management of GBL, evaluation of GBL, which are emerging to become key factors in developing the games literacy framework for teachers and continuing education for successful GBL.

4.4. Analytic Hierarchy Process (AHP) on Phase Three's email survey

The weight value of each evaluation index is the eigenvector value of the comparison matrix (see Table 2). The calculated weights of the two dimensions relative to the target layer are 0.5025 (Practical game-based teaching literacy) and 0.4975 (Generate games literacy). In turn, among the items contained in practical game-based teaching literacy, instructional design of GBL weights the largest (0.3348), followed by organisation and management of GBL (0.3337) and evaluation of GBL (0.3315). In general games literacy, the proportion of high-level games literacy (0.5028) is slightly higher than the basic games literacy (0.4972).

The results of AHP provided a partial answer to Research Question 2: What are the most important capabilities of games literacy that teachers need to prioritise? It indicated that instructional design for GBL and high-level games literacy were the literacy capabilities to be prioritised affecting the quality of game-based teaching utmost. The perceptions among the 20 experts of capability development for teachers' games literacy were collected and analysed in the method of AHP. Their perceptions were consolidated and became judgmental criteria to ascertain the most important capabilities that should be prioritised and more targeted for teachers to develop their games literacy. The result suggested that teachers need to prioritise instructional design GBL and high-level games literacy as guidelines for the improvement of their teaching practices in the GBL contexts.

Table 2. The weights and consistency test results

Target layer	Constructs	% of weights	Domains	% of weights
0.5025	Instructional design of GBL	0.3364	Game creation	0.1973
			Curriculum design of GBL	0.2032
			Activity design of GBL	0.1998
			Product design of GBL	0.1990
			Design of GBL environment	0.2002
	Organisation and management of GBL	0.3354	Gamification management of classes and schools.	0.5095
			Organisation & implementation of GBL	0.4950
	Evaluation of GBL	0.3332	Evaluate learning with gamification	0.2506
			Evaluate GBL	0.2491
			Game critique	0.2497
			Game belief	0.2506
0.4975	Basic games literacy	0.4972	Game knowledge	0.2491
			Game access	0.2497
			Game manipulation	0.2506
			Game morality	0.2506
	High-level games literacy	0.4978	Game spirits	0.3315
			Game thinking	0.3348
			Game identify	0.3337

5. Developing a Games Literacy for Teacher Education (GLTE) framework

Previous research on GBL has rarely focused on the capabilities in games literacy for teachers, and the measurable capabilities were much needed to be explored (Osatuyi et al., 2018). The present research evaluated experienced teachers' and researchers' perceptions of GBL. In order to explore the latent variables in games literacy, a synthesis of quantitative and qualitative research conducted was necessary and supported the development of an emerging framework for teachers to develop games literacy in GBL. This framework provides guidelines for instructors to integrate games into teaching and learning at all levels in different settings to implement GBL.

The proposed Games Literacy for Teacher Education (GLTE) framework illustrates the connections between design and delivery of GBL and teacher's capability development in games literacy. We present this framework drawing on the insights of experienced teachers and scholars to emphasise practical understandings of GBL and situate teachers' capabilities development as key for the successful practices in teaching and learning.

Our framework was inspired by a model for Critical Games Literacy (Apperley & Beavis, 2013) which they argued that digital games are different to other digital media due to players' enactment. This model focused on the learning opportunity and learner-centred approach which provided a good foundation to our framework for teachers and their continuative professional development; specifically, presented as two interlocking layers: games-as-action and games-as-text. "The games-as-action addresses the experience of gameplay by examining the virtual worlds of digital games and the dynamic interplay between game and player. The games-as-text examines the connection between the digital game and the lifeworld of the player, where the game play is embedded, enacted and given meaning" (Apperley & Beavis, 2013, p. 2).

This model is clearly developed from the learner's perspective and intended to provide a framework for game-based curriculum planning and pedagogy as it characterises the features of digital games and game play. However, it lacks the clear guidance for teachers to engage in this model. It mentions that design "embraces several crucial and related meanings and should be regarded as both noun and verb indicating the relationship between meaning making elements on a screen or page, and action – the process of designing as a creative activity, with multimodal literacy reconceptualised as 'design' " (Apperley & Beavis, 2013, p. 8).

In our proposed framework, we regard teachers as designers and facilitators whose capability in games literacy directly resulting in the quality of teaching and learning in the complex GBL contexts. Based on the research findings, five literacy capabilities for teachers' game literacy development can be used to measure the quality of teaching and construct a systematic process of design and delivery of GBL: (1) instructional design of GBL, (2) evaluation of GBL, (3) organisation and management of GBL, (4) basic games literacy, and (5) high-level games literacy.

5.1. Instructional design of GBL

The literacy of instructional design of GBL is highlighted by the experts in the related fields, regarded as the most important literacy weighted the highest in comparison with the equivalent constructs, containing 7 items (see Table 2) about GBL. It elaborates the importance of instructional design before adopting games in the classroom. Games design is essential for games literacy (Buckingham & Burn, 2007), as it closely relates to learning outcomes (Gauthier & Jenkinson, 2018). Notably, these correspond to other games literacy in this research, that supports the development of this framework.

It has to be admitted that the teachers involved in the study were working in urban areas, which means that the first-order barrier (extrinsic to teachers) and the second-order barrier (intrinsic to teachers) (Ertmer,1999) are not obviously high when teachers use technology in the class. However, many teachers are still confronted with the third-order barrier: design thinking (Tsai & Chai, 2012). The dynamics of creation might be an obstacle when encountering the advancement of GBL. The improvement of design thinking helps overcome both the first and second order barriers as presumably all barriers are seen as problems that need to be tackled and resolved through human creative thinking (Tsai & Chai, 2012; Makki, O'Neal, Cotten & Rikard, 2018). It is revealed that teachers should focus on the instructional design of GBL which is identified as a key literacy construct in games literacy.

5.2. Evaluation of GBL

The literacy construct of evaluation of GBL contains 6 items (see Table 2). These emphasise the instructors' attitudes toward adopting games in teaching as an effective educational approach, and the implementation of GBL depends upon the role of teachers to a large extent (Ma, 2018). Teachers should first examine games from a critical perspective taking pros and cons of games into consideration and accept this concept of teaching – potentially increase the willingness to incorporate games in the class.

A critical analysis of student motivation, social growth and cognitive gains is important for teachers to recognise effectiveness and accessible functionality of selective games that might be feasible for teaching, as a result, their belief can be shifted to acknowledging the advantages of games for learning (Prestridge, 2017). The evaluation of GBL centres the purpose of the analysis that is not to just represent an object for evaluation, more importantly an evaluation method to observe the process of learning (Chen & Zhang, 2019). Educators are expected to be cautious about evaluating existing empirical evidence before adopting any new pedagogy (such as games and game play) in their own practices.

5.3. Organisation and management of GBL

Although the organisation and management of GBL is not as critical as instructional design for GBL, it had been carefully considered in teaching practices, according to the research findings. It does matter because instructors are comfortable with using games endorsed elements in class such as badges, leader boards, social competition, and reward system. These are all conducive to encourage engagement and increase students' motivation (Kapp, 2012; Hamari et al., 2016).

On the other hand, as mentioned earlier, fears about using games or conducting GBL do exist because some teachers are not confident in dealing with unexpected events or requests while teaching a game-based class, and to some extent that they may be challenged to find ways to bring students back to focus on learning content towards the end of each game. Such fears are subtle and underpinning the competency of teachers in digital technology and pedagogical practices, which are also reflected on the underlying beliefs and the pedagogies that teachers might be used to appropriate digital games in the classroom (Prestridge, 2017). When teachers use games in the classroom, they need necessary operating skills to control the length of the process and avoid the unnecessary delays to the class schedule. Therefore, the organisation and management capabilities are specifically needed in teacher training, especially for the novice cohort.

5.4. Basic games literacy

Basic games literacy promotes GBL although it may not be limited to education. However, teachers are required to acquire such literacy skills as they should be aware of good games and understand the importance of games selection impacting on the successful GBL implementation (Becker, 2017). The acquisition of this literacy enables educators to access educational games and experience of gameplays at a degree of ease, which caters for the needs of young learners and reduce the personal resistance due to lacking support in digesting basic technological knowledge.

The result of the interviews in the study confirmed that it was difficult to design and carry on high-quality game-based activities if teachers were not aware of basic games literacy. Consequently, it could prevent instructors from experiencing and perceiving the essence of the games in which their students can interact and engage. Teachers who are less exposed to the concept of games are less likely to comprehend the intended purpose of games design and potentially fail to engage students with games for learning purposes. Games ethics are served as principles to design and adopt games. Thus, the basic games literacy is key in the implementation process of typical digital technologies because games cannot even roll out until all the necessary resources and elements are put in place. This factor could be regarded as necessary resources and capability to conquer mainly the first-order barrier before a teacher starts to embrace games into class.

5.5. High-level games literacy

Not only is the high-level games literacy (See Table 2) welcome instructors, but also embracing all gamers. According to the result of AHP, game thinking is the most important quality, which has been described as an umbrella term encompassing gamification, serious game, game-inspired design, and play that can be used for problem-solving (Armstrong, Landers & Collmus, 2016). The key to integral of games and education lies on the connection between knowledge and games, which are delivered to students through creative designs. This kind of design as indicated by Zimmerman (2009) requires a systems-based game thinking.

Game spirits value the characteristics such as self-contained, active and open, self-generational, self-renewal, non-subjective and dialogue. Integralism, one of Gadamer's three interpretations of the games, refers to the re-creative and recurring spiritual communication by a mutual involvement of the game entities (Gadamer, Weinsheimer & Marshall, 2004). It inspires us to explore a relationship between teaching and gaming from a unique perspective integrating the support of game spirits (Wang, 2002). Teachers explore the possibility of GBL through game spirits, with consistency and continuity, that can be observed in stimulating creativity and innovation since teachers are encouraged to be creative in instructional design as game spirits afford. The instructors' willingness to implement GBL is usually high, whereas actions carried by instructors are rare due to the existing problems such as the three-level barriers (Zhu & Wang, 2018). This indicates that instructors who are with game spirits tend to show more persistence and endurance when challenges occur in the procedures.

Game spirits also encourage instructors to make continuous efforts to improve games literacy to advancing their game-based learning pedagogies. Game identity aims to solve the confusion between the virtual avatars and the real humans. Once the game is introduced in the classroom, the role of an instructor changes – thus it is necessary for instructors to take ownership of games and claim their gaming identity, which will enhance the effectiveness of games for education.

The GLTE framework complemented the research findings and responded to Research Question 2. This framework urges teachers to take an opportunity of exercising gameplays to improve games literacy in the implementation of GBL; however, it is not to recommend teachers of sparing all their time on this activity, or becoming a gamer (Prestridge, 2017). The purpose of gameplays is to acknowledge the effort of games being designed for education and be aware of good games for learners that can be used for teaching. Prior to adopting a game in the classroom, instructors are expected to improve general games literacy capabilities including, understandings of the knowledge related to games and GBL, and highlighting different game types, rules and operation. This fosters a foundation for stimulating aspirations for better instructional design, ensuring that games are effectively played in the classroom and met students' expectation for learning through selecting meaningful games with joy.

6. Implications and conclusion

The present study focuses on the essential literacy capabilities of instructors in the field of GBL Its significance is inevitable as little research has touch based on this matter. This study aims to provide a guidance through developing a framework addressing key and prioritised capabilities that will enable teachers to develop games literacy to cater for learner's needs in the GBL contexts. This framework can be used for educators to evaluate their practices of GBL. It specifically highlights the significance of instructional design for teachers to uptake the GBL pedagogy. As Zimmerman (2009) argued, games literacy requires a growing conversation and attention in all sorts of educational settings, given the moving forward digital or online learning in the current events it is important to build a rigorous connection group within the faculty, school or organisation. Through developing this framework, we are also calling for more research-led practices involving both teachers and researchers to establish ongoing collaboration.

Five literacy constructs for educators were identified that can be further explored, confirmed, and updated. The importance of these capabilities in games literacy ranked by AHP will assist educators in identifying key features and examine individual capabilities more efficiently and precisely. The result of AHP showed that practical game-based teaching literacy is equally important to general games literacy, providing a practical suggestion for instructors – that is it is necessary to be cautious about developing games literacy as well as strengthening teaching practices. It is certain that developed games literacy helps instructors accumulate materials and trigger inspiration for game-based learning. In addition, it helps eliminate the games and gaming caused potential knowledge gaps between instructors and students.

This research strives to be rigorous and critical. However, limitations are worthy of consideration. In the aspect of verification of the validity of the first survey data, opportunities can be provided to conduct Confirmatory Factor Analysis. For future research, the framework can be verified by means of post-survey or -tests with other cohorts, including students. It is suggested that future research explores the relationship between these literacy factors through mediation effects. In-depth analysis on the various literacy constructs proposed in this study and framework could be conducted to identify the procedures of instructional design and evaluate effective games design and each would often significantly impact on learning outcomes and experience.

Acknowledgement

We thank all the participating teachers and experts who assisted and involved in the process of data collection. We would like to express our gratitude to Professor Nian-Shing Chen, without his encouragement and mentorship, this paper would have not been possible. Our thanks are also extended to the anonymous reviewers for their valuable comments and suggestion to help strengthen this paper.

References

Apperley, T., & Beavis, C. (2013). A model for critical games literacy. *E-learning and Digital Media, 10*(1), 1-12.

Atkins, M. J. (1993). Theories of learning and multimedia applications: An Overview. *Research papers in Education*, *8*(2), 251-271.

Armstrong, M. B., Landers, R. N., & Collmus, A. B. (2016). Gamifying recruitment, selection, training, and performance management: Game-thinking in human resource management. In *Emerging research and trends in gamification* (pp. 140-165). Hershey, PA: IGI Global.

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Statistical Psychology*, *3*(2), 77-85.

Beavis, C. (2013). Multiliteracies in the wild: Learning from computer games. In G. Merchant, J. Gillen, J. Marsh, & J. Davies (Eds.), *Virtual literacies: Interactive spaces for children and young people* (pp. 57 – 74). New York, NY: Routledge.

Beavis, C. (2017). Serious play: Literacy, learning and digital games. In C. Beavis, M. Dezuanni, & J. O'Mara (Eds.), *Serious play: Literacy, learning and digital games* (pp. 1-17). New York, NY: Routledge.

Beavis, C., & O'Mara J. A. (2010). Computer games – pushing at the boundaries of literacy. *Australian Journal of Language and Literacy, 33*(1), 65-76.

Becker, K. (2017). *Choosing and using digital game in the classroom: A Practical guide.* Cham, Switzerland: Springer.

Bernard, H. R. (2017). *Research methods in anthropology: Qualitative and quantitative approaches*. Oxford, UK: Rowman & Littlefield, AltaMira Press.

Bhushan, N., & Ria, K. (2004). *Strategic decision making: Applying the Analytic Hierarchy Process*. London, UK: Springer-Verlag London Limited.

Bourgonjon, J. (2014). The Meaning and relevance of video game literacy. *CLCWeb: Comparative Literature and Culture*, *16*(5), 8.

Buckingham, D., & Burn, A. (2007). Game literacy in theory and practice. *Journal of Educational Multimedia and Hypermedia, 16*(3), 323-349.

Cantor, J., & Wilson, B. J. (2003). Media and Violence: Intervention strategies for reducing aggression. *Media Psychology, 5*(4), 363-403. doi:10.1207/S1532785XMEP0504_03

Carey, D. M. (1993). Teacher roles and technology integration. *Computers in the Schools, 9*(2-3), 105-118. doi:10.1300/J025v09n02_10

Chen, S., & Zhang, S. J. (2019). Gaming evaluation design for mobile learning activities. *Software Guide (educational technology), 18*(3), 64-68.

Chuang, T. Y., Chen, N. S., Chen, M. P., Shen, C. Y., & Tsai, C. M. (2011). Digital game literacy: The Difference between parents and their children. In *5th European Conference on Game Based Learning, ECGBL 2011* (pp. 106-113). Reading, UK: Academic Publishing Limited.

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston, MA: Pearson Education.

Day, K. (1973). Some theoretical and practical implications of one approach to simulation games. *Programmed Learning and Educational Technology*, *10*(4), 235-238.

del Blanco, Á, Torrente, J., Marchiori, E. J., Martínez-Ortiz, I., Moreno-Ger, P., & Fernández-Manjón, B. (2012). A Framework for simplifying educator tasks related to the integration of games in the learning flow. *Educational Technology & Society*, *15*(4), 305–318.

Dick, W., Carey, L., & Carey, J. O. (2005). *The Systematic design of instruction* (5th ed). New York, NY: Addison-Wesley Educational Publishers Inc.

Ertmer, P. A. (1999). Addressing first-and second-order barriers to change: Strategies for technology integration. *Educational Technology Research and Development*, *47*(4), 47-61.

Gauthier, A., & Jenkinson, J. (2018). Designing productively negative experiences with serious game mechanics: Qualitative analysis of game-play and game design in a randomized trial. *Computers & Education*, *127*, 66-89.

Gee, J. P. (2003). *What video game have to teach us about learning and literacy.* New York, NY: PalGrave-McMillan.

Gadamer, H. G., Weinsheimer, J., & Marshall, D. G. (2004). *EPZ truth and method*. London, UK: Bloomsbury Publishing.

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior*, *54*, 170-179.

Hsu, C.-Y., Tsai, M.-J., Chang, Y.-H., & Liang, J.-C. (2017). Surveying in-service teachers' beliefs about game-based learning and perceptions of technological pedagogical and content knowledge of games. *Educational Technology & Society*, *20*(1), 134–143.

Hummel, M. (2001). *Supporting medical technology development with the Analytic Hierarchy Process*. Groningen, The Netherlands: Rijksuniversiteit Groningen.

Jackson, J. (2009). Game-based teaching: what educators can learn from videogames. *Teaching Education*, *20*(3), 291-304.

Kaiser, H. F. (1974). An Index of factorial simplicity. *Psychometrika, 39*(1), 31-36

Kali, Y., McKenney, S., & Sagy, O. (2015). Teachers as designers of technology enhanced learning. *Instructional Science, 43*, 173-179. doi:10.1007/s11251-014-9343-4

Kapp, K. M. (2012). *The gamification of learning and instruction*. San Francisco, CA: Wiley.

Kim, Y. E. (2008). Study on curriculum construction of game literacy. *Korean Journal of Journalism& Communication Studies*, *52*, 58-84.

Klimmt, C. (2009). Key dimensions of contemporary video game literacy: Towards a normative model of the competent digital gamer. *Eludamos Journal for Computer Game Culture,* (1)*,* 23-31.

Ma, Y. F. (2018, December). *An Empirical study on the level and attitude of teachers' game-based teaching*. Paper presented at the 3rd international Conference on Game Based Learning, ICGBL 2018 (Paper 3), Hong Kong, China.

Malone, T. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science, 4*, 333–369.

Mama, M., & Hennessy, S. (2013). Developing a typology of teacher beliefs and practices concerning classroom use of ICT. *Computers & Education*, *68*, 380-387.

Makki, T. W., O'Neal, L. J., Cotten, S. R., & Rikard, R. V. (2018). When first-order barriers are high: A Comparison of second-and third-order barriers to classroom computing integration. *Computers & Education*, *120*, 90-97.

McKenney, S. (2005). Technology for curriculum and teacher development: Software to help educators learn while designing teacher guides. *Journal of Research on Technology in Education, 38*(2), 167–190.

Merrin, W. (2009). Media studies 2.0: Upgrading and open-sourcing the discipline. *Interactions: Studies in Communication & Culture, 1*(1), 17-34. doi:10.1386/iscc.1.1.17_1

O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes: Evidence from the computer games literature. *The Curriculum Journal, 16*, 455–474.

Osatuyi, B., Osatuyi, T., & De La Rosa, R. (2018). Systematic review of gamification research in IS education: A Multi-method approach. *Communications of the Association for Information Systems*, *42*(1), 5.

Partington, A. (2010). Game literacy, gaming cultures and media education. *English Teaching: Practice and Critique*, *9*(1), 73-86.

Pivec, P. (2009). *Game-based learning or game-based teaching?* Becta. Retrieved from https://dera.ioe.ac.uk/1509/1/becta_2009_emergingtechnologies_games_report.pdf

Prestridge, S. (2017). The non-gamer teacher, the quiz and pop teacher and the kinect teacher. In C. Beavis, M. Dezuanni, & J. O'Mara (Eds.), *Serious play: literacy, learning and digital games* (pp. 87-101). New York, NY: Routledge.

Prestridge, S., & de Aldama, C. (2016). A Classification framework for exploring technology enabled practice - FramTEP. *Journal of Educational Computing Research, 54*(7), 901-921. doi:10.1177/0735633116636767

Robertson, A. (2018). *The Importance of video game literacy for healthy parenting.* Parenting for a Digital Future. Retrieved from https://blogs.lse.ac.uk/parenting4digitalfuture/2018/09/05/the-importance-of-video-game-literacy/

Rowan, L., & Beavis, C. (2017). Serious outcomes from serious play: Teachers' beliefs about assessment of game-based learning in schools. In C. Beavis, M. Dezuanni, & J. O'Mara (Eds.), *Serious play: literacy, learning and digital games* (pp. 169-185). New York, NY: Routledge.

Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences, 1*(1), 83-98. doi:10.1504/IJSSci.2008.01759

Salen, K. (2007). Gaming literacies: A Game design study in action. *Journal of Educational Multimedia and Hypermedia*, *16*(3), 301-322.

Shaffer, D. W., Squire, K. R., Halverson, R., & Gee, J. P. (2005). Video games and the future of learning. *Phi Delta Kappan*, *87*(2), 105-111.

Sitzmann, T. (2011). A Meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64*, 489–528.

Tao, K. (2009). On Reading Pictures "game literacy" and its elements. *Modern Distance Education Research,* (2), 14-18.

Tollånes, M. C., Aarsand, A. K., & Sandberg, S. (2011). Excess risk of adverse pregnancy outcomes in women with porphyria: A Population-based cohort study. *Journal of Inherited Metabolic Disease*, *34*(1), 217-223.

Wang, Y. M. (2010). Conceptual construction and framework research of game literacy in the perspective of new media. *China's Information Industry,* (10), 33-35.

Wang, Y. M. (2002). When technology meets beliefs: Preservice teachers' perception of the teacher's role in the classroom with computers. *Journal of Research on Technology in Education*, *35*(1), 150-161.

Wouters, P., & van Oostendorp, H. (2013). A Meta-analytic review of the role of instructional support in game-based learning. *Computers & Education, 60*, 412-425. doi:10.1016/j.compedu.2012.07.018

Yeh, Y. C., Chang, H. L., & Chen, S. Y. (2019). Mindful learning: A Mediator of mastery experience during digital creativity game-based learning among elementary school students. *Computers & Education, 132,* 63-75.

Yong, A. G., & Pearce, S. (2013). A Beginner's guide to factor analysis: Focusing on Exploratory Factor Analysis. *Tutorials in Quantitative Methods for Psychology, 9*(2), 79-94.

Yukselturk, E., Altıok, S., & Başer, Z. (2018). Using game-based learning with Kinect technology in foreign language education course. *Educational Technology & Society*, *21*(3), 159-173.

Zagal, J. P. (2008). A Framework for games literacy and understanding games. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share* (pp. 33-40). doi:10.1145/1496984.1496991

Zagalo, N. (2010). Creative game literacy: A Study of interactive media based on film literacy experience. *Comunicar*, *17*(35), 61-68.

Zhu, S. M., & Wang, T. (2018). Empirical research on gamification learning in basic education in China. *Digital education, (4),* 50-55.

Zimmerman, E. (2009). Gaming literacy: Game design as a model for literacy in the twenty-first century. In *The video game theory reader 2* (pp. 45-54). New York, NY: Routledge.

Appendix A. Semi-structured interview questions

1. Do you play games? How long/often do you spend on gameplays weekly? What types of games do you play?
2. How do you think gameplays? Do you think there is a relationship between playing games and using games for learning and teaching in the classroom?
3. If you were invited to give a lecture for teachers about GBL, what procedures would you follow?
4. Could you describe an impressive GBL experience you have encountered? Any challenging experience you have encountered?
5. What capabilities are required for teachers to design and deliver GBL?

Appendix B. Game Literacy in Teaching Practice (GLTP) survey

I. Demographic information
Gender: Male / Female
Grade taught: Kindergarten / Primary School / Middle School / High School / College
Teaching age: less than 1 year / 1-3years / 4-10 years / 10-20 years / over 20 years
Incumbent disciplines: Literature/Language/Math/Information technology/Physical science/Social science/Others
Weekly game time: less than 2hours/ 2-7hours / 7-14 hours /over 14 hours

II. Formal questionnaire
In the following items, please give your opinion by choosing 1 to 5 to express the importance of these literacy from "very unimportant" to the "very important."

I think teachers who conduct GBL should be of (following literacy):
D1: ability to combine the game-based teaching activities with teaching objectives and content
D2: ability to develop game products or tools based on teaching objectives and conditions
D3: ability to use game products or tools according to the teaching objectives and conditions
D4: ability to design and develop game-based teaching activities based on teaching objectives
D5: ability to design and develop game-based curriculum based on teaching objectives
D6: ability to use external resources to create a gamified physical environment and atmosphere
D7: ability to adapt gamification products or tools according to teaching objectives and actual conditions
M1: ability to apply gamification in the management of classes and schools.
M2: ability to organize and implement game-based teaching activities
M3: ability to organize and implement game-based courses
A1: ability to discover and evaluate a gamified school environment;
A2: ability to evaluate the game-based teaching process in a gamified manner (i.e. using game elements or game mechanics in non-gaming situations);
A3: ability to assess the game-based teaching activities and courses;
A4: ability to evaluate the pros and cons of a game software/application/tool

U1: knowledge to understand game elements (such as leaderboards, rewards, timely feedback, etc.) U2: knowledge to understands the concepts and concepts related to game-based teaching and Game-based learning; U3: ability to create the game which could be used in teaching; U4: skill to game operation; U5: ability to access a game, such as know how to download/find a game;
C1: ability to distinguish the pros and cons of a game objectively, C2: ability to distinguish the different among game genres; C3: ability to view the game from a critical point of view; C4: The belief that the game can be used to solve practical problems better; C5: ability to recognizes the virtual thing in the game. It does not use plug-in to hang up, flash back, compete with the rules of the game, and experience online game in a timely, appropriate and moderate manner; C6: ability to distinguish the virtual world in game and the real world; C7: awareness to realize that game can be reflected in real-world situations and may have a reference to real life.
H1: game think to solve practical problems by using game elements and game mechanics; H2: ability to distinguish the identity and role in the game world and the real world; H3: spirit to confront challenges, never give up, keep an optimistic mentality when playing games.

Appendix C. The average random consistency index table RI

The average random consistency index table RI. (The calculation results of 1000 positive reciprocal matrix)

Matrix order	1	2	3	4	5	6	7	8
RI.	0	0	0.52	0.89	1.12	1.26	1.36	1.41
Matrix order	9	10	11	12	13	14	15	
RI.	1.46	1.49	1.52	1.54	1.56	1.58	1.59	

Appendix D. Example for Analytic Hierarchy Survey scoring matrix

Game-based teaching product design VS *game-based learning curriculum design* indicates the importance of *i* compared with *j*; game-based teaching products refer to various teaching aids and learning tools for game-based teaching.	1	3	5	7	9	1/3	1/5	1/7	1/9	other
Curriculum design of GBL VS. Activity design of GBL										
Activity design of GBL VS. Product design of GBL										
Curriculum design of GBL VS. Product design of GBL										