Journal of

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

Journal of Educational Technology ୭ Society

The Journal of Educational Technology & Society has Impact Factor 3.522 and 5-Year impact factor 3.941 according to Thomson Scientific 2020 Journal Citations Report.

> vol.25 no.3

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

Educational Technology & Society

Volume 25 Issue 3

July 2022

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

Educational Technology & Society

An International Journal

Aims and Scope

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

ET&S has established a solid and stable editorial office where the Editors-in-Chief aims to promote innovative educational technology research based on empirical inquires to echo the pedagogical essentials of learning in the real world—lifelong learning, competency-orientation, and multimodal literacy in the 21st century.

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

Founding Editor

Kinshuk, University of North Texas, USA.

Journal Steering Board

Nian-Shing Chen, National Taiwan Normal University, Taiwan; Kinshuk, University of North Texas, USA; Demetrios G. Sampson, University of Piraeus, Greece.

Editors-in-Chief

Maiga Chang, Athabasca University, Canada; Andreas Harrer, Dortmund University of Applied Sciences and Arts, Germany; Yu-Ju Lan, National Taiwan Normal University, Taiwan; Yu-Fen Yang, National Yunlin University of Science and Technology, Taiwan.

Associate Editors

Yun Wen, National Institute of Education, Singapore

Editorial Board Members

Ahmed Hosny Saleh Metwally, Northeast Normal University, China; Bernardo Pereira Nunes, The Australian National University, Australia; Ching-sing Chai, The Chinese University of Hong Kong, Hong Kong; David Gibson, Curtin University, Australia; Grace Yue Qi, Massey University, New Zealand; Ig Ibert Bittencourt Santana Pinto, Universidade Federal de Alagoas, Brazil; Jerry Chih-Yuan Sun, National Chiao Tung University, Taiwan; Jie Chi Yang, National Central University, Taiwan; Joice Lee Otsuka, Federal University of São Carlos, Brazil; Kaushal Kumar Bhagat, Indian Institute of Technology, India; Minhong Wang, The University of Hong Kong; Morris Siu-Yung Jong, The Chinese University of Hong Kong; Regina Kaplan-Rakowski, University of North Texas, USA; Rita Kuo, New Mexico Tech, USA; Robert Li-Wei Hsu, National Kaohsiung University, Taiwan; Tony Liao, NOAA Earth System Research Laboratories, USA; Wen-Ta Tseng, National Taiwan University of Science and Technology, Taiwan; Yanjie Song, Education University of Hong Kong; Ahmed Tili, Smart Learning Institute of Beijing Normal University, China; Chia-Wen Tsai, Department of Information Management, Ming Chuan University, Taiwan; Hsueh Chu Rebecca Chen, Education University of Hong Kong; Jun (Chen) Scott Hsieh, Asia University, Taiwan; Katrin Saks, University of Tartu, Estonia; Nur Hamid, Universitas Islam Negeri Walisongo Semarang, Indonesia; Sheng-Shiang Tseng, Tamkang University, Taiwan.

Managing Editor

Sie Wai (Sylvia) Chew, National Taiwan Normal University, Taiwan; Phaik Imm (Alexis) Goh, National Yunlin University of Science and Technology, Taiwan.

Editorial Assistant

Kao Chia-Ling Gupta, The University of Hong Kong, China; Yen-Ting R. Lin, National Taiwan Normal University, Taiwan.

Technical Manager

Wei-Lun Chang, National Taiwan Normal University, Taiwan.

Executive Peer-Reviewers

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

Publisher

International Forum of Educational Technology & Society

Host

National Taiwan Normal University, Taiwan

Editorial Office

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

Supporting Organizations

University of North Texas, USA University of Piraeus, Greece National Yunlin University of Science and Technology, Taiwan National Sun Yat-sen University, Taiwan

Website Maintenance

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

Advertisements

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

Abstracting and Indexing

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

Guidelines for authors

Submissions are invited in the following categories:

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

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

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

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

Submission procedure

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

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

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

The preferred formats for submission are Word document, and not in any other word-processing or desktop-publishing formats. For figures, GIF and JPEG (JPG) are the preferred formats. Please place figures in the manuscript.

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

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

Educational Technology & Society

Volume 25 Number 3 2022

Table of Contents

Full Length Articles

Help Seeking from Peers in an Online Class: Roles of Students' Help-seeking Profiles and Epistemic Beliefs <i>Kun Huang and Victor Law</i>	1–14
Effects of Mobile-Assisted Language Learning on EFL/ESL Reading Comprehension Rui Li	15–29
Engineering Students' Readiness for Online Learning Amidst the COVID-19 Pandemic: Scale Validation and Lessons Learned from a Developing Country <i>Md Abdullah Al Mamun, Md Akbar Hossain, Sayedus Salehin, Md. Shahadat Hossain Khan and</i> <i>Mahbub Hasan</i>	30–45
The Impact of Teaching Simulation Using Student Chatbots with Different Attitudes on Preservice Teachers' Efficacy Donggil Song, Eun Young Oh and Hyeonmi Hong	46–59
How is One Plus One More than Two? The Interaction between Two Players in Online Co-Creativity Tasks <i>Ching-Lin Wu</i>	60–70
Editorial Note	
Blockchain in Smart Education Ching-Hsien Hsu, Amir H. Alavi and He Li	71–73
Special Issue Articles	
Exploring the Effect of Spatial Ability and Learning Achievement on Learning Effect in VR Assisted Learning Environment Chia-Chen Chen and Liang-Yu Chen	74–90
A Grounded Theory Study of the Psychological Distance in Online Education Ling Zhang, Jiang Ye and Jing-xuan Wang	91–104
Design and Development of a Blockchain-Based Secure Scoring Mechanism for Online Learning Cheng-Ting Tsai, Ja-Ling Wu, Yu-Tzu Lin and Martin KC. Yeh	105–121

iii

Help Seeking from Peers in an Online Class: Roles of Students' Helpseeking Profiles and Epistemic Beliefs

Kun Huang^{1*} and Victor Law²

¹University of Kentucky, USA // ²University of New Mexico, USA // k.huang@uky.edu // vlaw@unm.edu *Corresponding author

(Submitted May 14, 2021; Revised October 6, 2021; Accepted October 25, 2021)

ABSTRACT: Help seeking is a self-regulated learning strategy, and peer help is an important form of interaction in online education. Yet, students often do not seek help even at the cost of lower performance. To understand the factors behind online students' commitment to peer help, this study implemented a peer-help discussion forum in an online course and investigated the relationship among students' help-seeking profiles, epistemic beliefs (EB), and their actual participation in peer help. The findings revealed a significant relationship between students' EB and the number of response posts in the peer-help forum. Moreover, EB moderated the relationship between students' help-seeking profiles and the number of response posts. Theoretical and practical implications are drawn from the findings.

Keywords: Help seeking, Epistemic beliefs, Online education, Self-regulation, Peer help, Online discussion

1. Introduction

Learner interaction is a key mechanism to promote online learning (Garrison et al., 2000). Peer help, a particular form of learner interaction, constitutes an important self-regulated learning strategy (Karabenick, 2011; Pintrich & DeGroot, 1990) and allows learners to build a knowledge community (Greer et al., 2000). Peer help is particularly important in online education where an instructor is not physically available in the classroom to offer help. Yet, studies show that online students often do not seek help even at the cost of lower performance (Mahasneh et al., 2012). Given the benefits of help seeking (Richardson et al., 2012) and the prevalence of online education, it is important to understand the mechanism of online learners' help seeking in order to facilitate it. In a previous study that implemented a peer-help discussion forum in an online course, students perceived the forum to be highly helpful and participated far beyond the course requirement (Huang & Law, 2018). As a follow-up, the current study sought to understand the factors behind online learners' commitment to peer help, from the perspective of learners' epistemic beliefs (EB) and their help-seeking profiles such as help-seeking attitudes, tendencies, and goals. The findings would help develop a better understanding of online learners' help seeking and offer implications for encouraging online peer help. We start by reviewing relevant literature which led to the specific research questions we sought to answer.

2. Literature review

2.1. Help seeking

Help seeking is an effective self-regulated learning strategy to overcome academic difficulties (Karabenick, 2011; Pintrich & DeGroot, 1990). Nelson-Le Gall (1985) delineates five stages of help seeking: awareness of the need for help, deciding to seek help, identifying helpers, employing strategies to elicit help, and evaluating help. Throughout the stages, various learner-related factors play a role in whether, for what reason, and from whom learners seek help (Ryan et al., 2001). For instance, when help seeking threatens their self-esteem, learners may choose not to seek help despite the awareness of such a need (Karabenick & Knapp, 1991; Ryan et al., 2001). In such instances where learners need but do not seek help, they display a help-seeking avoidance (Ryan & Pintrich, 1997). Even when they decide to seek help, learners may opt for different sources; some seek formal help from teachers, while others turn to fellow students for help (Karabenick, 2003). Further, learners may aim for different goals in obtaining help (Nelson-Le Gall, 1985). Those aiming for executive goals count on the help to complete a task, e.g., getting a direct answer to a multiple-choice question. On the other hand, those with instrumental goals focus on getting just enough assistance to complete a task by themselves; instead of a direct answer, they may ask for a hint so that they can find out the answer on their own. Research has studied the relationships among the aforementioned learner factors in help seeking. It was found that help-seeking avoidance is negatively correlated with instrumental goals while positively correlated with help-seeking threat and executive goals (Karabenick, 2003; Pajares et al., 2004).

Traditionally, help seeking has been studied in the classroom setting. With an increasing role of modern technologies in education, researchers have started to examine help seeking in interactive learning environments (e.g., Aleven et al., 2003; Roll et al., 2014) and more recently, in online education. The myriad of help sources (e.g., instructors, peers, online resources) and channels (e.g., emails, synchronous meetings, discussion boards) renders a complex picture of help seeking in online education, which can be different from traditional classroom settings. For instance, it was found that online students feel less threatened in seeking help than those in the classroom (Kitsantas & Chow, 2007; Qayyum, 2018). Further, while students in classrooms are more likely to seek informal help from peers (Karabenick & Knapp, 1991), online students appear to have mixed preferences (Mahasneh et al., 2012; Makara & Karabenick, 2013; Qayyum, 2018).

A few research gaps exist in the current literature on help seeking. First, most research to date focuses on examining the relationship between learners' help-seeking profiles and learning outcomes. Richardson et al.'s (2012) meta-analysis suggested that help seeking has a small positive correlation with GPA. More recent literature further supported the connections between help seeking and academic performance (Martin-Arbos et al., 2021; Wu, 2021). A body of research generally highlights the positive effects of instrumental help from formal sources (e.g., instructors), as well as the negative effects of help-seeking threat, executive help seeking, and avoidance tendency (Karabenick, 2003; Kitsantas & Chow, 2007; Ryan & Pintrich, 1997; Schenke et al., 2015). What is less known is the connection between learners' help-seeking profiles and their engagement in help seeking, e.g., learners' behaviors in seeking help. Second, most research on help seeking is restricted to the face-to-face setting, which may not apply to the online environment (Bartholome et al., 2006; Er et al., 2015). For instance, in contrast to previous research, a study suggested the positive effect of informal help in online education (Goda et al., 2013); additionally, a more recent study on MOOC students found that those who were more inclined to seek help were less likely to do well in passing assessments (Kizilcec et al., 2017). In consideration of the research gaps, it was the intent of this study to examine the relationship between students' help-seeking profiles and their behaviors in seeking informal help from peers in an online class.

2.2. Online discussions as a source and a traceable record of peer help

Online discussion boards play a significant role in online education for a variety of purposes, including being a source of help. By interacting and discussing with peers, knowledge is constructed through peer scaffolding (Ge & Land, 2004; Hmelo-Silver et al., 2011), which could potentially promote both social and cognitive presence in an online class (Garrison et al., 2000). On the other hand, seeking help from peers on a discussion board may also pose threats to those who do not want to appear "dumb" before others (Kim et al., 2018).

It has been critiqued that most research on help seeking relies on learners' self-reports, which could deviate from actual help-seeking behaviors (Cross et al., 2017; Mahasneh et al., 2012). On the other hand, online discussion boards afford a means of tracking learners' posting behaviors as a manifestation of help seeking (Karabenick & Berger, 2013). Indeed, the recent rise of learning analytics has enabled researchers to examine a variety of factors in online education through online discussion analytics (Martínez et al., 2020; Wise et al., 2014; Xie & Huang, 2014). Examining computer science students' help seeking through their posting behaviors (question and response posts) on a discussion forum, Bull et al. (2001) found the forum to be helpful to all students, including those who posted questions, responded to posts, or merely read the posts. More recently, in comparing a regular discussion forum with one that allowed students to invite friends or experts to participate in problem-solving tasks, Chao et al. (2018) found that the latter led to more participation in help seeking in terms of questioning, subscribing, commenting, and viewing activities. Similarly, Huang and Law (2018) implemented a peer-help discussion forum in an online class and found a significant relationship between students' course grades and their number of help-giving, but not help-seeking, discussion posts.

What has not been examined is the relationship between learners' help-seeking profiles (attitudes, tendencies, and goals) and their help-seeking posting behaviors in online discussions. Moreover, such a relationship is likely to be influenced by learners' fundamental beliefs about the nature of knowledge and knowing, i.e., their EB. The next section introduces EB and its potential roles in help seeking.

2.3. Epistemic beliefs and their roles in help seeking

Hofer and Pintrich (1997) defined EB as "individuals' beliefs about the nature of knowledge and the process of knowing" (p. 117). They suggested two dimensions of EB: beliefs about the nature of knowledge (certainty and simplicity of knowledge), and beliefs about the nature of knowing (source of knowledge and justification of knowing). EB is often depicted as a continuum ranging from "naïve" to sophisticated (Greene et al., 2018;

Schommer, 1990). An individual with naïve EB may, for example, believe that knowledge only comes from authority and is unchanging. Evidence is clear that sophisticated EB is generally linked to deep cognitive engagement, productive study strategies, and positive learning outcomes (DeBacker & Crowson, 2006; Greene et al., 2018; Greene & Yu, 2016).

EB has been identified as an antecedent of self-regulation that affects learners' goal setting in a learning situation and their subsequent use of learning strategies (Muis, 2008). Specifically, Muis and Franco (2009) found that naïve EB predicted performance goal orientations, which, in turn, negatively predicted the effective use of study strategies. Although Muis and Franco (2009) did not incorporate help seeking as a learning strategy in their model, it is plausible that EB might play a similar role in help seeking which is a self-regulated learning strategy (Pintrich & deGroot, 1990). Concerning the possible relationship between EB and help seeking, Aleven et al. (2003) postulated that EB might affect learners' perceived values of, engagement in, and approaches to help seeking. For example, learners with naïve EB may overestimate their knowledge status, which may affect their awareness of the need for help and engagement in help seeking (Aleven et al., 2003). There is some empirical evidence supporting these arguments. For instance, Bartholome et al. (2006) studied learners' use of help features in an interactive learning environment and found that learners with more sophisticated EB made more use of the help features. Hao et al. (2016) found that EB predicted computer science students' help seeking from teachers or online resources, but not that from peers.

In addition to its potential impact on help-seeking behaviors, EB may also affect the relationship between learners' help-seeking profiles and help-seeking behaviors. Specifically, for learners who show an avoidance tendency to seek help, if they hold sophisticated EB, they are likely to adopt a mastery goal for learning (Winbert et al., 2019). As such, despite the avoidance tendency, they may still decide to seek help to develop a genuine understanding. On the other hand, for those who have a similar avoidance tendency but hold naïve EB, their EB may lead them to adopt performance goals (Muis & Franco, 2009; Winbert et al., 2019). Accordingly, to avoid showing incompetence in front of others, these learners may not seek help. Taken together, we postulate the potential role of EB in moderating the relationship between learners' help-seeking profiles and behaviors, which the current study set out to investigate.

2.4. The present study

In summary, the literature review above identified a few research gaps. First, most research on help seeking examined the relationship between help-seeking profiles and learning outcomes in the classroom setting. More research is needed to examine the connection between learners' help-seeking profiles and actual help-seeking behaviors in online education. Secondly, there is a lack of understanding regarding the roles of EB in help seeking: how it might affect learners' help-seeking behaviors and moderate the relationship between the behaviors and learners' help-seeking profiles. In closing the gaps, this study implemented a peer-help discussion forum in a completely online class and sought to answer the following research questions:

- What is the relationship between students' help-seeking profiles and their posting behaviors in the peer-help discussion forum?
- What is the relationship between students' EB and their posting behaviors in the peer-help discussion forum?
- Does students' EB moderate the relationship between their help-seeking profiles and posting behaviors in the peer-help discussion forum?

3. Method

3.1. Participants and context

Forty-nine students (30 females and 19 males) at a U.S. southern university voluntarily participated in the study. The participants were recruited from two online sections of a course offered by the university's program in information technology services. The participants' ages ranged between 21 and 52, including 40 seniors (82%), 7 juniors (14%), and 2 Master's students (4%). The reported ethnicities are as follows: White 34 (69%), African American 13 (27%), and other 2 (4%). The course focused on Microsoft Office Suite for business communications and operations. Throughout the semester, students worked on 19 projects that required them to apply software skills to business practices. Specific guidelines were provided for each project. To complete a project, students had to review course materials, understand the business needs in the project scenario, apply software skills, and troubleshoot any emerging issues in the process.

While students were individually accountable to complete all the projects, a peer-help discussion forum was set up on the course's Blackboard Learning Management System to encourage collaborative learning. Specifically, students were asked to use the forum to seek or offer help regarding their projects. With a minimum contribution of three total posts by the end of the semester, the students could earn three percent of the course grade. The modest requirement was designed to avoid the situation where students participated only for the sake of earning points. Students were informed that the instructor would generally respond to a request for help if others did not respond within 24 hours. In the actual implementation, the instructor sometimes responded earlier when she deemed necessary. The peer-help forum was piloted in the same course the previous year. Students provided positive feedback regarding its helpfulness, and their actual participation far exceeded the minimum requirement (Huang & Law, 2018).

3.2. Instruments

Students' help-seeking profiles were measured with a 13-item questionnaire of five subscales: instrumental help-seeking goal, executive help-seeking goal, formal versus informal help seeking, help-seeking threat, and avoidance of help seeking. The items were adapted from Karabenick (2003). Specifically, the references in the original items about getting help from "this class" were modified as getting help from the peer-help forum. For instance, the original item, "Even if the work was too hard to do on my own, I wouldn't ask for help with this class," was revised to "Even if the work was too hard to do on my own, I wouldn't ask for help on the forum." Students rated their agreement with each item on a 5-point scale, ranging from "1-Not at all true of me" to "5-Very true of me." Karabenick (2003) reported the subscales to be in the acceptable range of internal consistency. The instrument has been administered in studies on college students, and the relationships among the subscales were confirmed by significant correlations, which supported the convergent validity of the instrument (Finney et al., 2018; Karabenick, 2003).

Instruments	Subscales	Example items
Help-seeking	Instrumental help	If I have trouble understanding a project, I would post to the
profiles	seeking	Forum to see if someone could help me understand the general
		ideas
	Help-seeking threat	I would feel like a failure if I ask for help on the Forum
	Help-seeking avoidance	Even if the work was too hard to do on my own, I wouldn't ask
		for help on the Forum
	Formal vs informal help	In this class, the instructor would be a better source of help than
	seeking	the Forum
Epistemic	Certainty of knowledge	Principles in this field are unchanging
beliefs	Source of knowledge	If you read something in a textbook for this subject, you can be sure it is true
	Justification of knowing	Correct answers in this field are more a matter of opinion than
	sustineation of knowing	fact
	Attainability of truth	If scholars try hard enough, they can find the answers to almost anything

Table 1. Subscales and sample items for help-seeking profiles and epistemic beliefs

Students' EB was measured with the Discipline-Focused Epistemological Beliefs Questionnaire (Hofer, 2000), which is one of the most widely used instruments for EB. Eighteen items measured four dimensions: certainty of knowledge, source of knowledge, justification of knowing, and attainability of truth. Participants rated their agreement with each statement on a 5-point scale ranging from "1-Completely disagree" to "5-Completely agree." The items were scored in such a way that lower scores indicated more sophisticated beliefs. The instrument was found to have a reasonable reliability (Hofer, 2000; Muis et al., 2014). In Cazan's (2013) study on the validity of the instrument, exploratory factor analysis yielded the same four subscales, and significant correlations were found between the instrument and another EB inventory, which supported convergent validity of the instrument. Table 1 lists the subscales and sample items for each instrument. The complete instruments are presented in Appendix A.

3.3. Procedure

Participants were recruited from their respective classes following the regulations of the university's institutional review board. Each week throughout the semester, students reviewed learning materials and completed assigned

projects. They also used the peer-help discussion forum throughout the 15 weeks of the course to ask questions and give help. Towards the end of the semester, the participants completed a web-based survey. The survey had three parts: Part 1 recorded students' demographic information, Part 2 measured their help-seeking profiles, and Part 3 measured their EB.

3.4. Data collection and analysis

In addition to the data collected from the 3-part survey, students' participation data in the peer-help forum were also collected. Specifically, all the posts in the forum were collected and each students' numbers of initial posts and response posts were counted (Bull et al., 2001; Chao et al., 2018). Figure 1 illustrates an example initial post and its subsequent response posts. In initial posts, students were likely to ask questions, whereas response posts involved mostly student interactions surrounding the questions.

Figure 1. A screenshot of the peer-help forum showing initial and response posts (student names are masked)



Data analysis was performed in three steps. In the first step, the reliability of all the scales was calculated. All the scales had acceptable Cronbach *a* ranging from .62 to .97, except the subscale of executive help-seeking goal (Cronbach $\alpha = .16$), which was subsequently dropped from further analysis. The second step generated the descriptive and correlation statistics of the variables. In the last step, multiple regression analyses were conducted to answer the research questions.

4. Results

Table 2 presents the descriptive statistics of the variables. The correlations among the variables are presented in Table 3. On average, students posted 13.25 messages to the forum (including initial and response posts), which far exceeded the minimum requirement of three posts to receive the full credit.

Table 2. Descriptive statistics of help-seeking profiles, EB, and number of forum posts

Variable	Mean	SD	Min	Max
Help Seeking				
Instrumental help seeking	3.89	1.02	1.00	5.00
Help-seeking threat	1.46	.85	1.00	5.00
Help-seeking avoidance	1.44	.66	1.00	5.00
Formal vs. informal help seeking	2.82	.86	1.33	4.67
Epistemic beliefs				
Certainty of knowledge	2.98	.68	1.63	4.63
Source of knowledge	3.31	.72	1.50	5.00
Justification of knowing	3.24	.69	1.50	5.00
Attainability of truth	4.20	.76	2.00	5.00
Number of posts				
Initial posts	4.52	4.84	0.00	21.00
Response posts	8.73	10.27	0.00	64.00

Table 3. Correlations among help-seeking profiles, EB, and number of forum posts

Variable	1	2	3	4	5	6	7
1. Instrumental help seeking	1.00	41**	58**	28**	.10	.20	.16
2. Help-seeking threat		1.00	.71**	.19	.03	22	25
3. Help-seeking avoidance			1.00	.37**	.14	36*	40**
4. Formal vs. informal help seeking				1.00	.03	05	.12
5. Epistemic beliefs					1.00	.04	40**
6. Number of initial posts						1.00	.24
7. Number of response posts							1.00
NY NY 40 * 07 ** 04							

Note. N = 49; *p < .05, **p < .01.

To test the effects of students' help-seeking profiles and EB on their participation in peer help, a series of regressions were conducted. To do so, interaction terms were first created by multiplying EB and the help-seeking profile subscales (Cohen et al., 2003). Next, we mean standardized all the independent variables to prevent possible multicollinearity, especially among the interaction terms (Cohen et al., 2003). Further, multicollinearity diagnostic tests were conducted, and the results showed that multicollinearity was not a significant issue (VIF < 10; Hair et al., 2006). Graphical analyses were conducted to confirm the assumptions of linearity, normality, and homoscedasticity. Finally, the results also suggested independence of observations (1 < Durbin-Watson statistics < 3; Field, 2005).

Two separate stepwise regressions were conducted with the dependent variables being the number of initial posts and the number of response posts, respectively (Cohen et al., 2003). The first regression model tested the effects of students' help-seeking profiles and EB on their numbers of *initial* posts. Neither the main-effect nor the interaction-effect models were significant (p > .05), with the R^2 ranging from .143 to .231.

The second regression tested the effects of students' help-seeking profiles and EB on their numbers of *response* posts. The first model, which tested the main effects of help-seeking profiles on the number of response posts, was not significant, $R^2 = .189$, F(4, 44) = 2.567, p > .05 ($f^2 = .233$). In the second model, EB was added as an additional predictor, which was found to predict significantly over and above help-seeking profiles, $\Delta R^2 = .100$, $\Delta F(1, 43) = 6.023$, p < .05 ($f^2 = .406$). EB was a significant predictor: t(43) = -2.454, p < .05. In the third model, four interaction terms were added as additional predictors, i.e., the interactions between EB and the four help-seeking profile subscales (instrumental help seeking, help-seeking threat, help-seeking avoidance, and formal vs. informal help seeking). The model was found to predict significantly over and above the main-effect model, $\Delta R^2 = .235$, $\Delta F(4, 39) = 4.799$, p < .01 ($f^2 = 1.096$). Specifically, three significant predictors emerged: EB, t(38) = -2.159, p < .05; EB x Formal Help Seeking, t(38) = -4.101, p < .01; and EB x Help-Seeking Avoidance, t(38) = 2.287, p < .05. Results of the multiple regressions are shown in Table 4.

Table 4. Results of multiple linear regression analyses predicting number of response posts							
Variables	Model 1	Model 2	Model 3				
	Help-Seeking Profiles	Help-Seeking	Main and interaction				
	β (<i>t</i> -value)	Profiles & EB	effects				
		β (<i>t</i> -value)	β (<i>t</i> -value)				
Instru_HS	793 (456)	.132 (1.688)	396 (257)				
HS_Threat	.996 (.413)	1.217 (.097)	1.143 (.534)				
HS_Avoid	-8.392 (-2.357*)	-6.785 (-1.974)	-4.403 (-1.454)				
Formal_HS	3.958 (2.170*)	3.24 (1.76)	1.000 (.615)				
EB		-6.862 (-2.454*)	-6.506 (-2.159 [*])				
EB x Instru_HS			3.508 (1.031)				
EB x HS_Threat			-8.881 (-1.031)				
EB x HS_Avoid			18.810 (2.287*)				
EB x Formal _HS			-10.809 (-4.101**)				
F	2.567	3.493*	4.62**				
$\varDelta F$		6.023^{*}	4.799**				
Model adjusted R^2	.116	$.206^{*}$.413**				
ΔR^2		$.090^{*}$.207**				

Table 4 Posults of multiple linear regression analyses predicting number of response posts

Note. Instru_HS: Instrumental Help Seeking; HS_Threat: Help-Seeking Threat; HS_Avoid: Help-Seeking Avoidance; Formal_HS: Formal vs. Informal Help Seeking; EB: Epistemic Beliefs. N = 49; *p < .05, **p < .01.



Figure 2. The relationship between formal help seeking and number of response posts as a function of

Figure 2 illustrates the significant interaction between EB and formal help seeking. As shown, formal help seeking had a different impact on the number of response posts as a function of students' EB. For those with sophisticated EB, the more they perceived the instructor as a formal source of help, the more likely they were to respond to peers' posts. However, for those with naïve EB, the more they identified with the instructor as a formal source of help, the less likely they chose to respond.

Figure 3 illustrates the significant interaction between EB and help-seeking avoidance. Similarly, the helpseeking avoidance tendency had a different impact on students' response posts depending on their levels of EB. For those with sophisticated EB, the more they reported an avoidance tendency, the less likely they were to respond to others' posts. On the other hand, those with naïve EB showed an opposite trend - the more they reported an avoidance tendency, the more likely they would respond.



Figure 3. The relationship between help-seeking avoidance and number of response posts as a function of EB

5. Discussion

In the context of a peer-help discussion forum in an online course, this study set out to examine the relationship between students' help-seeking profiles and their actual participation analytics in peer help. Further, students' EB was also investigated regarding their impact on peer-help participation, and their potential moderating role in the relationship between help-seeking profiles and peer-help behaviors.

5.1. Lack of evidence: Direct impact of help-seeking profiles on peer-help behaviors

For the first research question regarding the relationship between students' help-seeking profiles and their actual participation in peer help, the current study found that the help-seeking profile alone did not influence students' peer-help posting behaviors. In other words, students' perceived help-seeking threat, preferred sources of help, avoidance tendencies, and instrumental help-seeking goals did not directly affect their participation behaviors in the peer-help discourse, which included both initial posts and responses to others.

Taking advantage of the learning analytics afforded by an online discussion forum to trace peer-help behaviors, this study was among the first to examine such a relationship in an online setting, yet our finding suggests a lack of direct relationship, which implies that learners' help-seeking profiles may not manifest in their peer-help participation. Comparing our findings with the previous research, we see an interesting misalignment. Empirical evidence from the past suggested a positive relationship between help-seeking profiles and academic performance (Karabenick, 2003; Kitsantas, 2007; Ryan & Pintrich, 1997; Schenke et al., 2015). These positive relationships may imply that students with adaptive help-seeking profiles indeed seek help, which, in turn, improves their performance. However, our study was not able to find evidence between help-seeking profiles and actual peer-help behaviors, which suggests that help-seeking behaviors may not be the direct mechanism for students with adaptive help-seeking profiles to achieve positive performance. Indeed, Kizilcec et al. (2017) study found a negative relationship between MOOC students' inclination to seek help and their academic performance.

5.2. Impact of EB on peer-help posting behaviors

Regarding the second research question about the impact of EB on students' peer-help posting behaviors, our study confirmed the impact: EB emerged as a significant predictor of the number of peer-help response posts. Specifically, those with more sophisticated EB posted significantly more responses than those with less sophisticated beliefs.

The finding aligns with Bartholome et al. (2006) which found that learners with more sophisticated EB made more use of the help features in an interactive learning environment. However, the finding is different from Hao et al. (2016), which did not find EB to predict students' use of peer help. The difference might be due to Hao's et al. (2016) use of students' self-reported participation data, in contrast to our study's examination of the actual participation.

The finding is not surprising, given that sophisticated EB is characterized by an identification with the constructive nature of knowledge and knowing (King & Kitchener, 2004). Although such beliefs were not pronounced in the students' number of initial posts (likely help seeking), these beliefs nonetheless positioned epistemologically sophisticated students to be more open in responding to peers' posts, thereby contributing to the collaborative inquiry. Considering our earlier study that reported a significant relationship between students' numbers of help-giving posts and course grades (Huang & Law, 2018), we reason that those with more sophisticated EB were more likely to respond to peers' posts, which might be associated with better performance.

This finding also expands the literature on EB by offering evidence for the impact of EB on learners' willingness to react or respond to peer-help discourses. The existing literature pointed to a significant correlation between EB and academic performance (Greene et al., 2018; Greene & Yu, 2016) and offered further evidence to suggest that the impact mechanism is through learners' use of study strategies such as elaboration, critical thinking, metacognitive self-regulation, and rehearsal (Muis & Franco, 2009). Our study suggests learners' participation in peer help as an additional mechanism in the learning process that might contribute to the impact of EB on academic performance. Our finding further implies that compared with help-seeking profiles, online learners' EB is more likely to have a direct impact on their participation in peer help. Thus, learners' EB should be an important factor to consider in the design and delivery of online education.

5.3. Moderating role of EB in help-seeking posting behaviors

Regarding the third research question about the moderating role of EB in the relationship between help-seeking profiles and peer-help posting behaviors, our findings suggest that two dimensions of help-seeking profiles, formal help seeking and avoidance tendency, were significantly related to peer-help responses (but not initial posts) through the moderation of EB. Thus, although help-seeking profiles did not directly affect peer-help posting behaviors, they did exert an indirect influence with the moderation of students' EB.

Specifically, depending on students' EB, the two dimensions of help-seeking profiles each had a differential impact on students' number of response posts in the peer-help forum. The first dimension, formal help seeking, indicates students' tendencies to seek help from formal sources such as the instructor. For students who held more naïve EB, the greater their tendency for formal help, the less likely they were to respond to peers' posts. These students might believe that knowledge was quite certain, and the instructor held the authority of knowledge. As such, they might have placed more value in getting the right solution than the knowledgebuilding inquiry process leading to the solution (Aleven et al., 2003). Comparatively, epistemologically more sophisticated students displayed a different trend. Despite a greater tendency to seek the instructor for help, these students were nonetheless more likely to respond to peers' posts. The finding appears contradictory on the surface. However, part of the reason might lie in the fact that in contrast to initial discussion posts that more likely involved one seeking help, response posts were more likely to involve one offering help or to joining a conversation. Thus, despite their perception of the instructor as a better source of help, these students' sophisticated EB might have led them to see the need to participate in the peer-help forum as an established norm for help seeking in the course. Our earlier study (Huang & Law, 2018) of the same class in the previous year may also shed some light into the findings. The students from the previous class indicated in interviews that they had helped peers in the forum to promote a sense of "student camaraderie" which they felt "often lacking in an online course" (Huang & Law, 2018). While we could not tie these students to their levels of EB, these opinions might represent the reason why epistemologically more sophisticated students chose to respond to peers' posts.

The second dimension, avoidance tendency, indicates one's tendency to avoid seeking help. For students who held more naïve EB, the greater their tendency to avoid help seeking, the more likely they would respond to peers in the forum. This finding can be explained in two aspects. First, as discussed earlier, response posts were less intended for help seeking. Second, past research found that naïve EB predicted performance goals (Muis & Franco, 2009; Winbert et al., 2019) and was associated with an overestimate of one's understanding (Schommer, 1990; Schommer et al., 1992). Taken together, it is likely that despite their avoidance of seeking help, students with naïve EB were more likely to demonstrate their understanding in the forum by responding to others' posts, which is an indication of the performance goal orientation. Comparatively, students with more sophisticated EB tended to aim for mastery learning rather than for impressing others with their performance (Winberg et al., 2019). For these students with an avoidance tendency, one possibility was that they had a more realistic estimate of their own understanding. As such, they were less likely to respond to others if they were unsure of their knowledge. Huang and Law (2018) also found that some students were not willing to offer help because they "do not want to mislead (others)," which might explain the students' reasoning for a lack of participation. Epistemologically sophisticated students who did not have an avoidance tendency, on the other hand, would not hesitate to post initial messages to seek help and subsequently join the conversations with peers, which led to more response posts.

6. Conclusion

This study advances our understanding of help seeking from peers as an informal source of help that has a positive impact in online learning (Goda et al., 2013; Wu, 2021). Building on past research that took advantage of online discussions to study various aspects of online learning (Martínez et al., 2020; Wise et al., 2014; Xie & Huang, 2014), the current study implemented online discussions to facilitate and investigate learners' actual peer-help behaviors. The study extended the existing research that primarily focused on learning outcomes and relied on learners' self-reported data. While there were attempts to examine the relationship between EB and help-seeking behaviors (Bartholome et al., 2006; Hao et al., 2016), our study represents the first attempt to simultaneously model the relationships among help-seeking profiles, EB, and online peer-help behaviors. The findings highlight the important role of learners' EB in online peer help, which warrants further investigation.

One perspective to understand help seeking is through the lens of self-regulation (Pintrich, 2000). While Muis and Franco (2009) advanced a self-regulated learning framework that delineated EB's role in learners' goal setting and subsequent use of self-regulation strategies, our study adds to the framework by examining help seeking as one of the self-regulation strategies and providing empirical evidence for the relationship between EB and help seeking.

Further, the different findings between the numbers of initial posts and response posts prompted us to consider help giving as the other side of help seeking since the initial posts were more likely to seek help, whereas response posts were more intended to offer help. In the traditional paradigm of education, the instructor is often the main source of help, and help seeking from the instructor is connected to positive learning outcomes (Karabenick, 2003; Kitsantas & Chow, 2007; Ryan & Pintrich, 1997; Schenke et al., 2015). Yet, with the constructivist paradigm, collaborative learning becomes more and more prevalent (Jeong et al., 2019). Collaborative learning necessitates both help seeking and help giving among peers, while our understanding of help giving is disproportionate to that of help seeking. With the affordance of online learning, it becomes easier to capture and examine both help seeking and help giving. More studies are needed to understand and theorize peer help giving, especially in online learning environments.

Our study yields practical implications for online education. Help seeking is an important self-regulation strategy, especially in online education (Karabenick, 2011; Pintrich & DeGroot, 1990). While learners' help-seeking profiles bear significant importance in help seeking, our findings suggest learners' EB as an important consideration if we are to encourage meaningful and productive peer help. To develop online learners' EB, learning activities should emphasize justification and reconciliation of objectivity and subjectivity (Cartiff et al., 2021). Online instructors should also communicate to students the importance of informal help and peer discourse, while fostering a climate for achieving mastery in learning. Epistemologically more sophisticated students should be encouraged to participate more in peer help for the benefit of building a supportive knowledge community, and guidance and mechanisms should be provided for these students to feel more comfortable to contribute. Finally, this study joins many other studies in suggesting the importance of EB in education, especially in online education (Huang et al., 2019; Greene et al., 2018; Greene & Yu, 2016). Actions should be taken to consider learners' EB in designing instruction. Meanwhile, promoting learners' EB should be a priority for educational research and practice.

This study has several limitations that future research should address. First, while the small sample size was sufficient with a medium to large effect size (the three regression models with $f^2 = .23$, .41, and 1.10; Cohen, 1992), it could present weaker statistical power for the regression analyses. Future studies may use a larger sample size that would also enable the examination of the subscales of EB. Second, while the peer-help forum was intended for help seeking among peers, the availability of the instructor on the forum could have made an undue influence on students' use of the forum. Third, the two instruments in this study used different pronouns (you and I), which may have caused confusions in responding to the questions. Fourth, future studies could include students' performance data, goal orientations, and peer-help access data for a more comprehensive understanding of peer help and its impact. Finally, future research should also examine how help-seeking profiles may affect help seeking from sources other than an established peer-help forum.

References

Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research*, 73(3), 277-320.

Bartholome, T., Stahl, E., Pieschl, S., & Bromme, R. (2006). What matters in help-seeking? A Study of help effectiveness and learner-related factors. *Computers in Human Behavior*, 22, 113-129.

Bull, S., Greer, J., McCalla, G., & Kettle, L. (2001). Help-seeking in an asynchronous help forum. In *Proceedings of Workshop on Help Provision and Help Seeking in Interactive Learning Environments*. International Conference on Artificial Intelligence in Education.

Cartiff, B. M., Duke, R., & Greene, J. (2021). The Effect of epistemic cognition interventions on academic achievement: A Meta-analysis. *Journal of Educational Psychology*, *113*(3), 477-498.

Cazan, A-M. (2013). Validity of Discipline Focused Epistemological Beliefs Questionnaire (DFEBQ) on a Romanian sample. *Procedia - Social and Behavioral Sciences*, 78, 713-717.

Chao, P. Y., Lai, K. R., Liu, C. C., & Lin, H. M. (2018). Strengthening social networks in online discussion forums to facilitate help seeking for solving problems. *Educational Technology & Society*, 21(4), 39–50.

Cohen, J. (1992). A Power primer. Psychological Bulletin, 112(1), 155-159. https://doi.org/10.1037/0033-2909.112.1.155

Cohen, J., Cohen, P., West, S.G., & Aiken, L. (2003). Applied multiple regression/correlation analysis for the behavioral sciences. Lawrence Erlbaum Associates.

Cross, S., Waters, Z., Kitto, K., & Zuccon, G. (2017). Classifying help seeking behaviour in online communities. In *Proceedings of LAK 2017*. https://doi.org/10.1145/3027385.3027442

DeBacker, T., & Crowson, M. (2006). Influences on cognitive engagement: Epistemological beliefs and need for closure. *British Journal of Educational Psychology*, 76, 535-551

Er, E., Kopcha, T. J., & Orey, M. (2015). Exploring college students' online help-seeking behavior in a flipped classroom with a web-based help-seeking tool. *Australasian Journal of Educational Technology*, *31*(5), 537–555.

Field, A. (2005). Discovering statistics using SPSS. Sage Publications.

Finney, S. J., Barry, C. L., Horst, S. J., & Johnston, M. M. (2018). Exploring profiles of academic help seeking: A Mixture modeling approach. *Learning and Individual Differences*, *61*, 158-171.

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. https://doi.org/10.1016/S1096-7516(00)00016-6

Ge, X., & Land, S. M. (2004). A Conceptual framework for scaffolding ill-structured problem-solving processes using question prompts and peer interactions. *Educational Technology Research and Development*, 52(2), 5–22.

Goda, Y., Yamada, M., Matsuda, T., Kato, H., Saito, Y., & Miyagawa, H. (2013). Effects of help seeking target types on completion rate and satisfaction in e-learning. In *Proceedings of INTED 2013* (pp. 1399–1403).

Greene, J. A., Cardiff, B. M., & Duke, R. F. (2018). A Meta-analytic review of the relationship between epistemic cognition and academic achievement. *Journal of Educational Psychology*, *110*(8), 1084-1111.

Greene, J. A., & Yu, S. B. (2016). Educating critical thinkers: The Role of epistemic cognition. *Policy Insights from the Behavioral and Brain Sciences*, 3(1), 45-53.

Greer, J., McCalla, G., Cooke, J., Collins, J., Kumar, V., Bishop, A., & Vassileva, J. (2000). Integrating cognitive tools for peer help: The Intelligent intranet peer help-desk project. In S. P. Lajoie (Ed.). *Computers as cognitive tools (Vol. 2): No more walls* (pp. 69-96). Erlbaum.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Pearson Prentice Hall.

Hao, Q., Wright, E., Barnes, B., & Branch, R. M. (2016). What are the most important predictors of computer science students' online help-seeking behaviors? *Computers in Human Behavior*, 62, 467-474.

Hmelo-Silver, C. E., Jordan, R., Liu, L., & Chernobilsky, E. (2011). Representational tools for understanding complex computer-supported collaborative learning environments. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing Interactions in CSCL: Methods, Approaches and Issues* (Vol. 12, pp. 83-106). Springer US.

Hofer, B. (2000). Dimensionality and disciplinary differences in personal epistemology. *Contemporary Educational Psychology*, 25(4), 378-405.

Hofer, B., & Pintrich, P. R. (1997). The Development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, *67*, 88–140.

Huang, K., & Law, V. (2018). Learners' engagement online in peer help. American Journal of Distance Education, 32(3), 177-189.

Huang, K., Law, V., & Lee, S. J. (2019). The Role of learners' epistemic beliefs in an online community of inquiry. *British Journal of Educational Technology*, *50*(4), 1882-1895.

Jeong, H., Hmelo-Silver, C. E., & Jo, K. (2019). Ten years of computer-supported collaborative learning: A Meta-analysis of CSCL in STEM education during 2005-2014. *Educational Research Review*, 28, 1-17.

Karabenick, S. A. (2003). Seeking help in large college classes: A Person-centered approach. *Contemporary Educational Psychology*, 28, 37-58.

Karabenick, S. A. (2011). Methodological and assessment issues in research on help seeking. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 267-281). Routledge.

Karabenick, S. A., & Berger, J. (2013). Help seeking as a self-regulated learning strategy. In H. Bembenutty, T. J. Cleary, & A. Kitsantas (Eds.). *Application of self-regulated learning across diverse disciplines: A tribute to Barry J. Zimmerman* (pp. 237-262). Information Age Publishing.

Karabenick, S. A., & Knapp, J. R. (1991). Relationship of academic help seeking to the use of learning strategies and other instrumental achievement behavior in college students. *Journal of Educational Psychology*, *83*, 221–230.

Kim, S., Zhang, K., & Park, D. (2018). Don't want to look dumb? The role of theories of intelligence and humanlike features in online help seeking. *Psychological Science*, *29*(2), 171-180.

King, P., & Kitchener, K. (2004). Reflective judgment: Theory and research on the development of epistemic assumptions through adulthood. *Educational Psychologist*, *39*(1), 5–18.

Kitsantas, A., & Chow, A. (2007). College students' perceived threat and preference for seeking help in traditional, distributed, and distance learning environments. *Computer Science Education*, 48(3), 383–395.

Kizilcec, R. F., Perez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in massive open online courses. *Computers & Education*, 104, 18-33.

Mahasneh, R. A., Sowan, A., & Nassar, Y. H. (2012). Academic help-seeking in online and face-to-face learning environments. *E-Learning and Digital Media*, 9(2), 196–210. https://doi.org/10.2304/elea.2012.9.2.196

Makara, K. A., & Karabenick, S. A. (2013). Characterizing sources of academic help in the age of expanding education technology: A New conceptual framework. In S. A. Karabenick & M. Puustinen (Eds.), *Advances in help-seeking research and applications: The role of emerging technologies* (pp. 37–72). Information Age Publishing.

Martín-Arbós, S., Castarlenas, E., & Dueñas, J. (2021). Help-seeking in an academic context: A Systematic review. *Sustainability*, *13*. https://doi.org/10.3390/su13084460

Martínez, J. P. C., Catasús, M. G., & Fontanillas, T. R. (2020). Impact of using learning analytics in asynchronous online discussions in higher education. *International Journal of Educational Technology in Higher Education*, 17. https://doi.org/10.1186/s41239-020-00217-y

Muis, K. R. (2008). The Role of epistemic beliefs in self-regulated learning. Educational Psychologist, 42(3), 173-190.

Muis, K. R., Duffy, M. C., Trevors, G., Ranellucci, J., & Foy, M. (2014). What were they thinking? Using cognitive interviewing to examine the validity of self-reported epistemic beliefs. *International Educational Research*, 2(1), 17-32.

Muis, K. R., & Franco, G. M. (2009). Epistemic beliefs: Setting the standards for self-regulated learning. *Contemporary Educational Psychology*, 34, 306-318.

Nelson-Le Gall, S. (1985). Help-seeking behavior in learning. Review of Research in Education, 12, 55-90.

Pajares, F., Cheong, Y. F., & Oberman, P. (2004). Psychometric analysis of computer science help-seeking scales. *Educational and Psychological Measurement*, 64(3), 496-513.

Pintrich, P. R. (2000). The Role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451-502). Academic Press.

Pintrich, P. R., & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33–40.

Qayyum, A. (2018). Student help-seeking attitudes and behaviors in a digital era. *International Journal of Educational Technology in Higher Education*, 15(17). https://doi.org/10.1186/s41239-018-0100-7

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A Systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353-387.

Roll, I., Baker, R. S. J. d., Aleven, V., & Koedinger, K. R. (2014). On the benefits of seeking (and avoiding) help in online program-solving environments. *The Journal of the Learning Sciences*, 23, 537-560.

Ryan, A. M., & Pintrich, P. R. (1997). "Should I ask for help?" The Role of motivation and attitudes in adolescents' help seeking in math class. *Journal of Educational Psychology*, 89(2), 329-341.

Ryan, A. M., Pintrich, P. R., & Midgley, C. (2001). Avoiding seeking help in the classroom: Who and why? *Educational Psychology Review*, *13*(2), 93–114. https://doi.org/10.1023/A:1009013420053

Schenke, K., Lam, A. C., Conley, A. M., & Karabenick, S. A. (2015). Adolescents' help seeking in mathematics classrooms: Relations between achievement and perceived classroom environmental influences over one school year. *Contemporary Educational Psychology*, *41*, 133-146.

Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology*, 82, 498–504.

Schommer, M., Crouse, A., & Rhodes, N. (1992). Epistemological beliefs and mathematical text comprehension: Believing it is simple does not make it so. *Journal of Educational Psychology*, *84*, 435–443.

Winberg, T. M., Hofverberg, A., & Lindfors, M. (2019). Relationships between epistemic beliefs and achievement goals: Developmental trends over grades 5-11. *European Journal of Psychology of Education*, 34, 295-315.

Wise, A. F., Zhao, Y., & Hausknecht, S. N. (2014). Learning analytics for online discussions: Embedded and extracted approaches. *Journal of Learning Analytics*, 1(2), 48-71.

Wu, J. -Y. (2021). Learning analytics on structured and unstructured heterogeneous data sources: Perspectives from procrastination, help-seeking, and machine-learning defined cognitive engagement. *Computers & Education, 163.* https://doi.org/10.1016/j.compedu.2020.104066

Xie, K., & Huang, K. (2014). The Role of beliefs and motivation in asynchronous online learning in college-level classes. *Journal of Educational Computing Research*, 50(3), 317-343.

Appendix A

Help seeking scales

- 1. If I were having trouble understanding a project, I would post to the Forum to see if someone could help me understand the general ideas.
- 2. Getting help from the Forum would be one of the first things I would do if I were having trouble with a project.
- 3. The purpose of seeking help from the Forum would be to succeed without having to work as hard.
- 4. Getting help from the Forum would be a way of avoiding doing some of the work.
- 5. I would feel like a failure if I ask for help on the Forum.
- 6. I would not ask for help on the forum to let others find out that I needed help.
- 7. Seeking help on the Forum would be an admission that I am just not smart enough to do the work on my own.
- 8. If I didn't understand something in this class I would guess rather than asking for help on the Forum.
- 9. Even if the work was too hard to do on my own, I wouldn't ask for help on the Forum.
- 10. I would rather do worse on an assignment I had trouble with than seeking help from the Forum.
- 11. If I were to seek help in this class I would directly ask the instructor rather than posting to the Forum.
- 12. I would prefer seeking help on the forum rather than directly asking the instructor. (rev)
- 13. In this class, the instructor would be better to get help from than would the Forum.

Discipline Focused Epistemological Beliefs Questionnaire

- 1. Truth is unchanging in this subject.
- 2. In this subject, most work has only one right answer.
- 3. Sometimes you just have to accept answers from the experts in this field, even if you don't understand them.
- 4. All professors in this field would probably come up with the same answers to questions in this field.
- 5. If you read something in a textbook for this subject, you can be sure it is true.
- 6. Most of what is true in this subject is already known.
- 7. In this subject, it is good to question the ideas presented. (rev)
- 8. Correct answers in this field are more a matter of opinion than fact.
- 9. If scholars try hard enough, they can find the answers to almost anything.
- 10. Experts in this field can ultimately get to the truth.
- 11. Principles in this field are unchanging.
- 12. If my personal experience conflicts with ideas in the textbook, the book is probably right.
- 13. There is really no way to determine whether someone has the right answer in this field.
- 14. Answers to questions in this field change as experts gather more information. (rev)
- 15. All experts in this field understand the field in the same way.
- 16. I am more likely to accept the ideas of someone with first-hand experience than the ideas of researchers in this field.
- 17. I am most confident that I know something when I know what the experts think.
- 18. First-hand experience is the best way of knowing something in this field.

Effects of Mobile-Assisted Language Learning on EFL/ESL Reading Comprehension

Rui Li

School of Foreign Languages, Hunan University, Hunan, China // liruidianzi@hotmail.com

(Submitted July 2, 2021; Revised September 5, 2021; Accepted October 25, 2021)

ABSTRACT: While an increasing number of studies have cast light on the effectiveness of MALL (mobileassisted language learning) on English as a foreign/second language (EFL/ESL) reading comprehension, there is still a lack of comprehensive meta-analysis regarding the effect sizes of these studies. To fill the gap, this study reported results based on a meta-analysis of 20 effect sizes from 17 experimental and quasi-experimental studies published during 2000–2020. The results showed that the overall effect size was significantly large, suggesting the use of MALL applications for EFL/ESL reading comprehension is more effective than traditional methods. The moderating effects of eight moderators were analyzed. The intervention settings and intervention durations were found to be significant moderators, while others did not find a significant moderating effect. Implications of the findings were discussed.

Keywords: Meta-analysis, Mobile-assisted language learning (MALL), Reading comprehension

1. Introduction

Compared with other language skills, the development of reading ability is the foundation of foreign/second language (FL/L2) learning (Li, 2021a). Motivated by the possibility of creating portable, connective, context-sensitive, location-aware, multifunctional and ubiquitous learning environments, the use of mobile-assisted language learning (MALL) applications, e.g., smartphones, tablets and e-readers, has to date been proven to be useful to develop L2 learners' language skills (Burston, 2014; Burston, 2015; Hwang & Fu, 2019; Li & Hafner, 2022; Shadiev et al., 2020), L2 reading comprehension in particular (e.g., Gutiérrez-Colón et al., 2020; Klimova & Zamborova, 2020; Lin et al., 2020; Moon et al., 2021). Research into the use of MALL for L2 reading comprehension is crucial because a better understanding of the effects and the related moderators will inform reading pedagogy.

While researchers (e.g., Keezhatta & Omar, 2019; Mays et al., 2020; Sofiana & Mubarok, 2020) have paid much attention to empirically examine the pedagogical affordances of MALL for L2 reading comprehension, and numerous studies have obtained the facilitative effects in "increasing reading frequency, allowing for peer interaction, achieving higher sustained attention and acquiring better decoding skills" (Lin et al., 2020, p. 851), synthesized empirical evidence of its impact on reading comprehension is still lacking. Compared with the empirical studies, meta-analysis results are more reliable and generalizable, as they are based on results of multiple studies and increased sample sizes (Yanagisawa et al., 2020). In order to narrow the gap, this study synthesized various literature on MALL for L2 reading comprehension and conducted a meta-analysis to provide a more up-to-date vision on this issue. Specifically, it aims to (a) empirically generalize findings of previous MALL for L2 reading research while (b) dealing with the variability of the aggregated effects from a meta-analytic perspective.

2. Literature review

2.1. Related studies on MALL for L2 reading comprehension

Reading is an active, dynamic and complex cognitive process that involves the selection of relevant information, the mapping of information into a mental representation and the integration of information using existing knowledge (Lin et al., 2020). Previous studies (e.g., Davis & Lyman-Hager, 1997; Singhal, 1998; Whitford & Joanisse, 2018) have examined readers' decoding and metacognitive processes, L2 lexical accessibility, L2 reading strategies, cultural differences and L2 subskills of reading comprehension. In recent decades, the introduction of MALL technologies, such as cellphones (Chen et al., 2011; Sofiana & Mubarok, 2020), tablet PCs (Lin, 2017; Lan et al., 2013) and PDAs (Hsu et al., 2013; Wu et al., 2011), has reshaped the traditional L2 reading pedagogical paradigm, and researchers were having mixed and inconclusive findings towards the shift.

On the one hand, some researchers (Alemi & Lari, 2012; Lin, 2014; Mays et al., 2020) have found the facilitative effects of MALL for L2 reading comprehension. For instance, Alemi and Lari (2012) adopted a quasi-experiment to investigate the effect of vocabulary learning with SMS (short messaging service) on L2 reading comprehension. Results indicated that the experimental group outperformed the control group in L2 reading performance. In another quasi-experiment, Mays and colleagues (2020) explored the use of mobile ARS (Audience Response Systems) with student-generated questioning on EFL learners' reading comprehension. The results indicated that the quality of questions provided by the experimental group improved at a greater rate over time compared with the control group. Participants of the experimental group also have a higher level of collaboration and engagement than those of the control group.

On the other hand, other researchers (Chen et al., 2011; Lin, 2017) have obtained the limited effects of MALL for L2 reading comprehension. For instance, Chen and colleagues (2011) conducted a quasi-experiment to compare the effectiveness of direct access to digital materials with QR (quick response) codes and that of scaffolded questioning in improving EFL learners' reading comprehension. Results suggest that the MALL technology did not influence EFL learners' reading comprehension, rather the traditional approach with scaffolded questioning improved their reading performance. Similarly, Lin (2017) also conducted a quasi-experiment to examine the effectiveness of a MALL technology on EFL learners' reading performance. Results did not find any significant difference between both groups.

Taken together, although the aforementioned studies have been helpful in shedding some light on the use of MALL for L2 reading comprehension, the discrepancy among them might be explained by a number of moderators, such as proficiency levels, educational levels, screen sizes, software types, intervention settings, intervention durations, instructional approaches and measured outcome types, according to the existing studies (Gutiérrez-Colón et al., 2020; Sung et al., 2015). The present study, therefore, was promoted by a need to revisit the moderators that may moderate the effects of MALL for L2 reading comprehension.

2.2. Related reviews of MALL for L2 reading comprehension

To date, several reviews (Gutiérrez-Colón et al., 2020; Klimova & Zamborova, 2020; Lin et al., 2020; Reiber-Kuijpers et al., 2021) on MALL for L2 reading comprehension have been conducted. For instance, Gutiérrez-Colón and colleagues (2020) provided a thorough review of MALL for L2 reading research between 2012 to 2017. Results of their study indicated that future study should focus on the use of appropriate mobile device types, the use of appropriate screen sizes for mobile reading and the application of mobile device in the informal settings. Lin and colleagues (2020) reviewed the literature on MALL for L2 reading comprehension during 2008 to 2018. Based on the review results, they provided design-related, strategy-related and learner-related guidelines. In another review, Klimova and Zamborova (2020) conducted a literature search of 21 articles on MALL for L2 reading comprehension and coded them based on research objective, participants, MALL technology, intervention durations, outcomes, main results and limitations. In a more recent review, Reiber-Kuijpers and colleagues (2021) systematically synthesized digital reading in FL/L2 in relation to digital reading environments, tasks, readers, and strategy use during 2008 to 2020. Results of their review suggest that future attempt should be extended to informal settings, and researchers should explore reading in more authentic environments and consider the important role of teachers. Although these studies have afforded insights into trends of MALL for L2 reading comprehension, studies published to date have neither directly calculated the aggregated effects of MALL for L2 reading comprehension, nor dealt with the variability of the aggregated effects with moderator analyses, which suggests an urgent need to meta-analyze the effects of MALL for L2 reading comprehension and examine whether the calculated effects were moderated by a series of moderators, including proficiency levels, educational levels, screen sizes, software types, intervention settings, intervention durations, instructional approaches and measured outcome types.

2.3. Research purposes and questions

Two research purposes should be achieved regarding the effects of MALL for L2 reading comprehension and its related moderators. First, drawing on the data collected from the primary studies, a meta-analysis was conducted on the aggregated overall effect sizes of MALL for L2 reading comprehension. It should be pointed out here that the experimental group used MALL for L2 reading comprehension, e.g., PDAs, e-reader, tablet PCs and cellphones, while the control group used non-MALL for L2 reading comprehension, including traditional pencil and paper (Lin, 2017; Lan et al., 2013), traditional teacher-centered lectures (Priyanti et al., 2019; Wang, 2017), PowerPoint (Mays et al., 2020), paper-based materials (Wu et al., 2011), traditional classroom instructions (Wu et al., 2010) and non-personalized instructions (Hsu et al., 2013), among others. Second, informed by several

existing meta-analyses (Chen et al., 2020; Sung et al., 2015), the moderating effects were conducted in relation to proficiency levels, educational levels, screen sizes, software types, intervention settings, intervention durations, instructional approaches and measured outcome types. Consequently, two research questions to be addressed are as follows.

- Research question 1 (RQ1): What is the overall effect size of MALL for L2 reading comprehension vs. non-MALL for L2 reading comprehension?
- Research question 2 (RQ2): How do moderators, such as proficiency levels, educational levels, screen sizes, software types, intervention settings, intervention durations, instructional approaches and measured outcome types, affect the aggregated effect sizes?

3. Research design

3.1. Literature retrieval

We adopted a careful and exhaustive literature retrieval approach to investigate the effectiveness of MALL applications on EFL/ESL learners' reading development. Drawing on the insights of recently published reviews (Gutiérrez-Colón et al., 2020; Klimova & Zamborova, 2020; Lin et al., 2020), the potential keywords and/or keyword combinations used in those review articles were consulted to promote a comprehensive search. Related primary studies were searched from several electronic online databases (e.g., web of science, ScienceDirect, Springer, ProQuest, Scopus, Wiley, ERIC) and search engines (Google Scholar and Baidu Scholar) by using a combination of the following MALL-related and reading-related keywords integrated with Boolean operators. The following Boolean expressions of keywords, i.e., (mobile-assisted language learning OR MALL OR mobile applications OR portal devices OR handheld devices OR mobile technologies OR mobile learning OR mlearning OR ubiquitous learning OR u-learning OR mobile phones OR cellphones OR smartphones OR e-reader OR tablets OR personal digital assistants OR PDAs OR gamification) AND (reading OR reading competence OR reading skills OR reading comprehension OR reading abilities OR reading performance), were executed. Second, to further avoid the insufficient search of a significant portion of the relevant literature in the first-round, we conducted a second-round backward and forward citation search based on the review articles (Gutiérrez-Colón et al., 2020; Klimova & Zamborova, 2020; Lin et al., 2020), along with snowballing technique (Biernacki & Waldorf, 1981) by scanning references in the identified articles (e.g., Gheytasi et al., 2015; Keezhatta & Omar, 2019; Naderi & Akrami, 2018). Third, informed by Sung and colleagues (2015), we also manually searched publications in the following major CALL journals (e.g., Computer Assisted Language Learning, Language Learning & Technology, ReCALL, System and CALICO Journal) and educational technology journals (e.g., Educational Technology & Society, Computers & Education, Internet and Higher Education, Computers in Human Behavior, British Journal of Educational Technology, Educational Technology Research and Development, Journal of Computing in Higher Education, Journal of Educational Computing Research, Journal of Computer Assisted Learning, Australian Journal of Educational Technology, Interactive Learning Environments, and The Asia-Pacific Education Researcher, among others) to further avoid the incomplete inclusion.

3.2. Inclusion and exclusion criteria

A total of 81 studies pertinent to MALL for L2 reading comprehension were identified via the initial literature retrieval. The following inclusion and/or exclusion criteria were proposed to ensure whether the retrieved studies were eligible for the meta-analysis. In what follows, a second-round manual inclusion and/or exclusion was executed.

- (1) Publications that were written in English should be confined to 2000–2020. This time range was chosen because MALL technologies remained few in number before 2000 (Duman et al., 2014). Second, to obtain a more comprehensive view, we intend to expand the time span of the recently published narrative reviews (e.g., ranging from 2008–2018 in Lin et al., 2020; from 2018–2020 in Klimova & Zamborova, 2020; and from 2012–2017 in Gutiérrez-Colón et al., 2020).
- (2) The study should adopt a form of MALL technologies (e.g., mobile phones, PDAs, computer tablets or ereaders) for EFL/ESL reading comprehension. Those studies that failed to use technologies or used MALL technologies on first or other foreign language (not EFL or ESL) reading comprehension were excluded.
- (3) The publications should contain sufficient statistics for data calculation or transformation of aggregated overall effect sizes. As such, only the experimental or quasi-experimental studies that examined the effectiveness of a mobile-assisted device on foreign or second language reading comprehension should be

included. More specifically, the independent variables should include different interventional modes (e.g., traditional learning method as a control group vs. mobile-assisted learning method as a treatment group), and the dependent variable should include a measure of the researcher-designed or standardized pre- and post-tests on reading comprehension between different modes. Those publications that investigated L2 learners' attitudes or perceptions, pedagogical or theoretical recommendations regarding mobile L2 reading comprehension were excluded. Furthermore, other review publications (e.g., review articles, book reviews, and editorial materials) were excluded as well. As a result, 21 eligible publications were finalized for meta-analysis.

3.3. Variables coded for each study

A well-designed coding scheme should "capture the pertinent information suitable for meta-analysis, including the identification of the publications, the characteristics of the participants, theoretically relevant features of the study and measured variables" (Wilson, 2019, p. 154). Explicit coding scheme was thus proposed in Table 1.

After the code scheme was developed, coding procedures were observed as follows: First, given the recommended practice for data dependencies (Plonsky, 2011; Plonsky & Oswald, 2014), multiple effect sizes reported in a single publication involved different participants or different types of measurement were coded separately to ensure the reliability of the analyses, resulting in a total of 21 eligible primary studies that yielded 24 effect sizes as independent studies. For instance, Hsu and colleagues (2013) conducted a quasi-experimental design that compared two experimental groups and one control group. The effect sizes were coded separately, since EFL learners of the experimental groups that involved different participants used a mobile language management system, while those of the control group adopted the traditional approach. Further, two experienced coders who are senior CALL researchers skilled at meta-analysis coding protocols negotiated with each other to ensure the consistent understanding of each variable and item. Then, they independently coded the items, and the interrater reliability (percentage of agreement) was 97%. The discrepancies were resolved by consensus through discussions.

Coding types	Subtypes	Operational definitions	References
Proficiency	1. low	Studies that reported learners as foreign	Li (2021a)
levels		language beginners.	
	2. intermediate	Studies that reported learners as intermediate	
		learners.	
Educational levels	1. primary education	Kindergarten or primary school students.	Li (2022)
	2. secondary	Junior middle school or senior high school	
	education	students.	
	3. tertiary education	College students.	
Screen sizes	1. small	Mobile phones and handheld PDAs.	Researcher-designed
	2. larger	Tablet PCs.	
Software	1. general purposes	Applications that were NOT designed for	Chen et al. (2020)
types		educational purposes, e.g., WeChat,	
		WhatsApp, telegram and QR codes.	
	2. educational	Applications that were designed for	
	purposes	educational purposes, e.g., language	
		management system.	
Intervention	1. formal/classroom	Formal learning activities that occurred in the	Chen et al. (2020)
settings		classroom.	
	2. informal/outdoor	Informal outside-of-the-classroom learning	
		activities.	
Intervention	1. one session, ≤ 1	Durations fewer than one week or only one	Chen et al. (2020)
durations	week	session.	
	2. >1 week, ≤4 weeks	Durations over one week, but fewer than four	
		weeks.	
	3. >4 weeks, ≤one	Durations over four weeks, but within one	
	semester	semester.	
Durations	Number of weeks	Studies that reported the specific number of	Xu et al. (2019)
_		weeks regarding intervention durations.	
Instructional	1. drill and practice	Studies that used MALL to practice L2	Researcher-designed

Table 1. The descriptive information of coding scheme

approaches		reading ability.	
	2. personalized	Studies that used MALL to appropriately	
	learning	provide reading materials to learners based on	
		their reading abilities.	
	3. game-based	Studies that adopted games as scaffolds to	
	learning	facilitate L2 reading.	
	4. multimedia	Studies that adopted multimedia resources or	
	learning	glosses to facilitate L2 reading.	
	5. collaborative	Studies that supported peer-peer and/or tutor-	
	learning	learner feedbacks or collaborations.	
Measured outcome types	1.standardized	Standardized TOEIC/IELTS/OPT reading comprehension test scores.	Li (2021a)
	2.researcher-designed	Reading comprehension test scores made by	
		researchers.	

Note. L2 = second language; TOEIC = Test of English for International Communication; IELTS = International English Language Testing System; OPT = Oxford Placement Test.

3.4. Calculation and analysis of the effect sizes

As the small sample sizes might bias the aggregated effect sizes, Hedges' *g* that "provided a simple correct for the bias" was taken as the effect size index for the eligible studies (Lipsey & Wilson, 2001, p. 48). When the original data reported in the primary studies did not include means and standard deviations, we used other statistical values, e.g., *t*-value, to calculate effect sizes. For instance, we used pre- and post-test means, sample size, and *t*-value to compute the effect sizes in several eligible studies (e.g., Ataee et al., 2015; Chen & Hsu, 3008; Gheytasi et al., 2015; Lan et al., 2013). The interpretations of the magnitude of an effect size were based on Plonsky and Oswald (2014): 0.200, 0.500, and 0.800 for small, moderate, and large effects, respectively.

4. Results

4.1. Outlier diagnosis results

According to Lipsey and Wilson (2001, p. 108), potential outliers with the extreme effect sizes that were "more than 3 standard deviations from the mean of all the effect sizes" should be excluded from the analysis. In doing so, four studies (g = 4.243, Grami & Hashemian, 2017; g = 4.632, Hazaea & Alzubi, 2016; g = 8.371, Keezhatta & Omar, 2019; g = 3.175, Motallebzadeh & Ganjali, 2011) out of 21 eligible studies that yielded extremely large effect sizes were excluded, resulting in a total of 17 remaining primary studies that yielded 20 independent studies (effect sizes) for the final analysis.

4.2. Publication bias analysis results

Since researchers normally did not publish nonsignificant results, the publication bias refers to the phenomenon that unpublished studies might differ from the published studies (Borenstein et al., 2009). The results of publication bias were often inspected via funnel plot and a fail-safe N method (Li, 2021a). As studies were distributed symmetrically (Figure 1), the probability of having a publication bias is rare. In addition, by evaluating how many unpublished studies with nonsignificant results would change the meta-analytic results from significant to nonsignificant, Rosenthal (1991) proposed a classical fail-safe N method to avoid the file-drawer problem. It was found that there existed no publication bias, since the result of fail-safe N was 950, which was significantly higher than the respective observed number 20 (z = 13.644, p < .001), that Rosenthal (1991) suggested for the file-drawer problem.

Figure 1. Funnel plot of the selected studies **Funnel Plot of Standard Error by Hedges's g**



4.3. Overall analysis results

The overall effect size was estimated using a random effect model, which "assumes that each observed effect size differs from the sampling error plus a value that represents other sources of variability" (Lipsey & Wilson, 2001, p. 119). As shown in Figure 2, the aggregated effect size computed from the 20 independent studies is large, g = 0.813, 95% CI = [0.566, 1.060] and significant, z(19) = 6.449, p < .001, indicating a positive and large effect for the use of MALL applications for L2 reading comprehension.

			Fig	ure 2.	Fores	st plo	t of the	e selec	ted studie	s			
Study name	Time poin	t	S	tatistics for	or each	study				Hed	ges's g and 95	% CI	
		Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value					
Ataee, Fatemi, & Ashraf	2015	0.766	0.214	0.046	0.347	1.185	3.580	0.000	1	1	12		\rightarrow
Chen & Hsu	2008	1.526	0.370	0.137	0.800	2.252	4.119	0.000					\rightarrow
Chen, Teng, Lee, & Kinshuk	2011	0.068	0.314	0.099	-0.548	0.683	0.215	0.830					
Gheytasi, Azizifar, & Gowha	n2015	2.161	0.425	0.180	1.329	2.994	5.088	0.000					>
Hsu, Hwang, & Chang_a	2013	0.649	0.236	0.056	0.186	1.112	2.749	0.006			-		>
Hsu, Hwang, & Chang_b	2013	0.552	0.235	0.055	0.093	1.012	2.354	0.019					>
Khubyari & Narafshan	2016	0.966	0.328	0.108	0.323	1.610	2.944	0.003					
Lan, Sung, & Chang_a	2009	0.383	0.276	0.076	-0.157	0.924	1.390	0.164					_
Lan, Sung, & Chang_b	2013	0.853	0.266	0.071	0.331	1.375	3.204	0.001				_	\rightarrow
Lin_a	2014	1.199	0.235	0.055	0.738	1.660	5.099	0.000					>
Lin_b	2017	0.036	0.206	0.042	-0.368	0.439	0.173	0.863				_	
Mays, Yeh, & Chen	2020	0.289	0.285	0.081	-0.271	0.848	1.012	0.312					- 1
Naderi & Akrami	2018	0.735	0.203	0.041	0.338	1.132	3.628	0.000					\rightarrow
Priyanti, Santosa, & Dewi	2019	0.581	0.237	0.056	0.118	1.045	2.458	0.014					>
Sofiana & Mubarok	2020	0.399	0.184	0.034	0.039	0.760	2.170	0.030					
Wang_a	2017	0.344	0.244	0.059	-0.134	0.821	1.412	0.158			_		- 1
Wang_b	2017	1.636	0.289	0.083	1.070	2.202	5.667	0.000					>
Wang_c	2017	0.577	0.249	0.062	0.090	1.064	2.322	0.020					>
Wu et ala	2011	2.534	0.304	0.093	1.938	3.131	8.327	0.000					>
Wu, Sung, Huang, & Yang	2010	0.734	0.236	0.056	0.271	1.198	3.105	0.002		1			
		0.813	0.126	0.016	0.566	1.060	6.449	0.000		1	- 1		
									-1.00	-0.50	0.00	0.50	1.00

Favours A Favours B

4.4. Homogeneity analysis

As apparent in Table 2, *Q*-value was 92.713 with p < .001, indicating that there were between-group differences among the effect sizes resulting from factors other than subject-level sampling error. The I^2 for the overall model

showed high heterogeneity ($l^2 = 79.507$), indicating that one or more moderators could account for this heterogeneity (Borenstein et al., 2005; Borenstein et al., 2009).

Table 2. Heterogeneity analysis results							
Q	df	р	I^2				
92.713	19	.000	79.507				

4.5. Moderator analysis results

The moderator analysis was carried out to examine the moderating effect of eight moderators, including proficiency levels, educational levels, screen sizes, software types, intervention settings, intervention durations, instructional approaches and measured outcome types. The moderator analysis results have been summarized in Table 3. Two moderators were found to have a moderating effect on the overall effect sizes: intervention settings and intervention durations, whereas the other moderators did not find a significant moderating effect.

Table 3. Moderator analysis results							
Moderators	k	g	z	95% CI	${\it Q}$ between	р	
Proficiency levels					1.586	0.208	
1. low	11	0.659	4.993***	[0.400, 0.917]			
2. intermediate	9	0.998	5.003***	[0.538, 1.457]			
Educational levels					2.461	0.292	
1. primary education	3	0.519	2.938**	[0.173, 0.866]			
2. secondary education	9	0.740	4.710***	[0.432, 1.048]			
3. tertiary education	8	1.003	3.848***	[0.492, 1.514]			
Screen sizes					0.048	0.826	
1. small	12	0.792	5.048***	[0.485, 1.100]			
2. larger	8	0.853	3.777***	[0.410, 1.295]			
Software types					0.080	0.778	
1. educational	15	0.794	5.280***	[0.499, 1.089]			
2. general	5	0.877	3.492***	[0.385, 1.369]			
Intervention settings					4.201^{*}	0.040	
1. formal/classroom	17	0.667	6.348***	[0.461, 0.873]			
2. informal/outdoor	3	1.684	3.471**	[0.733, 2.635]			
Intervention durations					6.626^{*}	0.036	
1. one session, ≤ 1 week	3	0.397	3.014**	[0.139, 0.656]			
$2. >1$ week, ≤ 4 weeks	4	0.948	2.342^{*}	[0.155, 1.742]			
3. >4 weeks, ≤one semester	13	0.895	5.752^{***}	[0.590, 1.200]			
Instructional approaches					2.145	0.709	
1. drill and practice	3	1.079	2.890^{**}	[0.347, 1.811]			
2. personalized learning	4	0.785	4.537***	[0.446, 1.124]			
3. game-based learning	2	0.786	1.966^{*}	[0.002, 1.569]			
4. multimedia learning	7	0.869	2.607^{**}	[0.216, 1.522]			
5. collaborative learning	4	0.601	4.671***	[0.349, 0.853]			
Measured outcome types					0.219	0.640	
1. standardized	8	0.734	3.545***	[0.328, 1.139]			
2. researcher-designed	12	0.868	5.348***	[0.550, 1.186]			
	. 1*	.1 1 1	c .1				

Note. k is the number of independent studies available for the certain variable; Hedges' g is effect size; CI is short for confidence interval; p < .05; p < .01; p < .001.

4.5.1. Proficiency levels

Proficiency levels included two categories: low (k = 11, 55%) and intermediate (k = 9, 45%). According to Table 3, intermediate proficiency learners achieved large effect size (g = 0.998, 95% CI = [0.538, 1.457]), and low proficiency learners obtained a moderate effect size (g = 0.659, 95% CI = [0.400, 0.917]). However, there was no statistical difference between the two proficiency levels, $Q_{\text{between}} = 1.586, p = 0.208$.

4.5.2. Educational levels

Three categories involved in educational levels: primary education (including pre-school, kindergarten and primary school education, k = 3, 15%), secondary education (junior middle school or senior high school education, k = 9, 45%) and tertiary education (college education and beyond, k = 8, 40%). As indicated in Table 3, EFL learners of tertiary education had large effect size (g = 1.003, 95% CI = [0.492, 1.514]), larger than those of the primary (g = 0.519, 95% CI = [0.173, 0.866]) and secondary education (g = 0.740, 95% CI = [0.432, 1.048]) that reported moderate effect sizes. However, no significant difference was found among the three educational levels, Q between = 2.461, p = .292.

4.5.3. Screen sizes

Screen sizes contained two categories: small (normally display sized from 3.5 to 7'', e.g., handheld cellphones and PDAs, k = 12, 60%) and larger (normally sized from 7 to 10.5'', e.g., tablets, k = 8, 40%). According to Table 3, compared with the small screen size (g = 0.792, 95% CI = [0.485, 1.100]), MALL applications with larger screen size (g = 0.853, 95% CI = [0.410, 1.295]) achieved much higher moderating effect. However, the difference did not reach a significance level, $Q_{\text{between}} = 0.048, p = .826$.

4.5.4. Software types

Software types included educational purposes (k = 15, 75%) and general purposes (k = 5, 25%). Table 3 showed using different types of MALL software resulted in moderate-to-large effect sizes. The effect size of educational purposes is 0.794, 95% CI = [0.499, 1.089], slightly lower than that of general purposes (g = 0.877, 95% CI = [0.385, 1.369]). Between-group comparison indicated no statistically significant difference, Q between = 0.080, p = .778.

4.5.5. Intervention settings

Two categories of intervention settings were involved: formal/classroom (k = 17, 85%) and informal/outdoor (k = 3, 15%). According to Table 3, significantly larger effect size was reported for using MALL applications in informal/outdoor setting (g = 1.684, 95% CI = [0.733, 2.635]) than that in formal/classroom setting (g = 0.667, 95% CI = [0.461, 0.873]), Q between = 4.201, p = .040.

4.5.6. Intervention durations

Intervention durations that were divided into three categories: "one session, ≤ 1 week" (k = 3, 15%), ">1 week, ≤ 4 weeks" (k = 4, 20%) and ">4 weeks, $\leq \text{one semester}$ " (k = 13, 65%) were found to have a significant moderating effect on the overall effect size, Q between = 6.626, p = 0.036. Post-hoc comparison was computed to locate the source of the moderator effect. "One session, ≤ 1 week" (g = 0.397, 95% CI = [0.139, 0.656]) had the weakest moderating effect, which was statistically lower as compared ">4 weeks, $\leq \text{one semester}$ " (g = 0.895, 95% CI = [0.590, 1.299]), Q between = 5.950, p = .015. Neither significant difference existed between ">1 week, ≤ 4 weeks" and "one session, ≤ 1 week" (Q between = 1.674, p = .196), nor between ">1 week, ≤ 4 weeks" and ">4 weeks, $\leq \text{one semester}$ " (Q between = 0.015, p = .902), indicating the longer the intervention durations, the larger the effect sizes.

4.5.7. Instructional approaches

Instructional approaches could be categorized into five approaches: drill and practice (k = 3, 15%), personalized learning (k = 4, 20%), game-based learning (k = 2, 10%), multimedia learning (k = 7, 35%) and collaborative learning (k = 4, 20%). Table 3 indicated drill and practice (g = 1.079, 95% CI = [0.347, 1.811]) and multimedia learning (g = 0.869, 95% CI = [0.216, 1.522]) had significantly high effect sizes, while game-based learning (g = 0.786, 95% CI = [0.002, 1.569]), personalized learning (g = 0.785, 95% CI = [0.446, 1.124]) and collaborative learning (g = 0.601, 95% CI = [0.349, 0.853]) had significantly moderate effect sizes. No between-group difference was observed, Q between = 2.145, p = .709.

4.5.8. Measured outcome types

Two measured outcome types could be categorized: standardized test (k = 8, 40%) and researcher-designed test (k = 12, 60%). Table 3 indicated no significant difference was found, $Q_{\text{between}} = 0.219$, p = .640, with the moderate-to-large effect size of standardized test being 0.734, 95% CI = [0.328, 1.139], and researcher-designed test being 0.868, 95% CI = [0.550, 1.186].

5. Discussion

The present study endeavored to quantitatively meta-analyze the overall effect size of MALL for L2 reading comprehension identified in the primary literature. Simultaneously, moderator analyses were also conducted to examine the moderating effects of proficiency levels, educational levels, screen sizes, software types, intervention settings, intervention durations, instructional approaches and measured outcome types for the effect size. The meta-analytical findings regarding two RQs were discussed in the remainder of this section.

5.1. Overall effect size of MALL for L2 reading comprehension

RO1 dealt with the overall effect size of MALL for L2 reading comprehension vs. non-MALL for L2 reading comprehension. A total of 20 effect sizes generated an overall aggregated effect size of 0.813 (95% CI = [0.566, 0.813]1.060]). The meta-analysis result indicated a positive and large effect for the use of MALL applications for L2 reading comprehension, suggesting that the use of MALL for L2 reading comprehension is more effective than traditional methods without MALL applications for L2 reading comprehension. A more informative interpretation of the results could be achieved by comparing the effect sizes with similar meta-analyses on MALL (e.g., Chen et al., 2020; Cho et al., 2018; Lin & Lin, 2019; Sung et al., 2015). The effect size (ES =0.813) of this study is much larger than those meta-analyses on MALL for language learning (Chen et al., 2020; Cho et al., 2018; Sung et al., 2015), but comparable with the results reported in a meta-analysis conducted by Lin and Lin (2019). Among those similar studies, Sung and colleagues (2015) conducted a meta-analysis of 44 MALL studies published between 1993 and 2003 on mobile devices for language learning, and found a moderate effect size of 0.55 for the use of MALL in language learning. Likewise, Cho and colleagues (2018) metaanalyzed 20 MALL studies published between 2005 and 2017, presenting a similar overall effect size of 0.51. In a more recent study, Chen et al. (2020) performed a meta-analysis of MALL in language learning based on 84 studies published during 2008 to 2018, and obtained a moderate-to-large effect size of 0.722. A plausible explanation for the discrepancy might be due to the different domains of investigation: domain-general vs. domain-specific. In other words, different from those domain-general meta-analyses (Chen et al., 2020; Cho et al., 2018; Sung et al., 2015) that dealt with MALL for language learning in general, the current study metaanalyzed the use of MALL for L2 reading comprehension in a domain-specific way. This explanation also lends support in another domain-specific meta-analysis (Lin & Lin, 2019), which systematically synthesized findings from 33 eligible studies published during 2005 to 2018, and also obtained a large effect size of 1.005 regarding MALL applications for L2 vocabulary learning.

5.2. Findings from testing for moderators

The moderating effects of proficiency levels, educational levels, screen sizes, software types, intervention settings, intervention durations, instructional approaches and measured outcome types were considered in RQ2.

Intervention settings. There were two settings involved: informal/outdoor setting vs. formal/classroom setting. In this study, significantly larger effect size was reported for using MALL applications in informal/outdoor setting (g = 1.684) than that in formal/classroom setting (g = 0.667), which suggests that learners who used MALL applications for informal L2 reading would outperform those did formally. This result is partly consistent with the findings of existing MALL studies (Chen et al., 2020; Sung et al., 2015), which claimed the stronger effect of learning with MALL applications in informal/outdoor setting than in formal/classroom setting. Intriguingly, while significant moderating effect of MALL applications for L2 reading comprehension was found in the current study, significant between-group differences were not reported in both studies that meta-analyzed MALL applications for language learning. Reasons might be that, unlike other domain-general language learning activities, the domain-specific MALL for L2 reading comprehension emphasizes more on the "increased self-practice outside class, independent learning and self-paced learning" (Lin et al., 2020, p. 853). Chen and Lin (2016) also asserted that EFL learners prefer reading in a self-paced informal way. In other words, "mobile

readers infrequently spend long time in formal reading, but they frequently spend short time in reading utilizing their spare time" (Chen & Lin, 2016, p. 568). Meanwhile, it should also be cautioned here that, the result of higher effect size for the informal/outdoor settings over formal/classroom settings did not mean formal classroom learning was not important at all. Rather, the optimal learning outcome could be achieved only when the formal classroom learning could be complemented by the informal outside-of-the-classroom learning (Sung et al., 2015).

Intervention durations. The moderator analysis results tended to support longer durations, especially the intermediate-term durations (">1 week, ≤ 4 weeks"). More specifically, the lowest effect size was found for short-term durations (ES = 0.397, "one session, ≤ 1 week"), followed by long-term durations (ES = 0.895, ">4 weeks, \leq one semester") and intermediate-term (ES = 0.948, ">1 week, \leq 4 weeks"). Although there was no significant difference between the short-term and intermediate-term, effect size of the long-term (ES = 0.895) was found to be significantly higher than that of short-term (ES = 0.397), indicating that long-term durations were favored, which is in alignment with the existing studies (Chen et al., 2020; Sung et al., 2015; Sung et al., 2016). The lowest effect for short-term suggests that the effectiveness of MALL for L2 reading comprehension might be limited, as learners need more time to be acquainted with using MALL applications and with the learning scenarios (Sung et al., 2015). Furthermore, the moderating effect of intermediate-term durations was slightly larger, though nonsignificant, than that of long-term durations suggests learners normally experienced novelty effect at the earlier stage of study (within the first 4 weeks) due to the curiosity and freshness of the MALL applications, and their sustained attention and the novelty effect would wear off for long-term of investigation (Chen et al., 2020). It should be cautioned when interpreting the statistical results and comparing the difference between the short-term (k = 3) and intermediate-term durations (k = 4) because of small numbers of studies involved, warranting further research in this regard.

Proficiency levels. There were two proficiency levels: low vs. intermediate. Moderator analysis of proficiency level showed intermediate proficiency learners achieved a large effect size of 0.998, and low proficiency learners obtained a moderate effect size of 0.659. This is consistent with Droop and Verhoeven (2003), which maintained that L2 reading involves a complex process of word decoding skills, morphosyntactic knowledge and intercultural awareness. It is thus understandable to speculate that intermediate proficiency learners would outperform low proficiency learners when using MALL for L2 reading comprehension, since "limited language proficiency has also been found to impede the L2 reading comprehension" (Droop & Verhoeven, 2003, p. 81).

Educational levels. Educational levels were roughly divided into primary, secondary and tertiary education. The largest effect size was found for tertiary education (ES = 1.003), followed by secondary education (ES = 0.740) and primary education (ES = 0.519), indicating the effect size increased from primary and secondary education to tertiary education, echoing Chen and colleagues (2020). One possible reason why learners of tertiary education received the largest effect size and learners of primary education had the smallest effect size was due to age effect. In other words, compared with young children and adolescents, learners of tertiary education are mature adults whose word decoding skills, morphosyntactic knowledge and intercultural awareness would be more mature to facilitate their L2 reading comprehension (Droop & Verhoeven, 2003).

Screen sizes. Screen sizes were roughly categorized into small screens (cellphones and PDAs) and larger screens (e.g., tablet PCs). It was found that effect size of larger screens was larger than that of small screens, which is well-attested in a number of studies (e.g., Chen & Lin, 2016; Gutiérrez-Colón et al., 2020; Wang & Higgins, 2005). For instance, Wang and Higgins (2005) found that the small screens on mobile devices would limit the amount and type of information that can be displayed. In the same vein, Gutiérrez-Colón and colleagues (2020) also noted that small screens may restrict how learners perceive the texts to be read, information transmission and attitude towards reading, so they could have limited access to reading a text in depth with high cognitive load, and, therefore, their reading performance on mobile devices may be negatively affected.

Software types. According to Chen et al. (2020), software types were classified into educational purposes and general purposes. The effect of using general-purpose applications is found to be slightly larger than that of educational-purpose applications. This result is contrary to Chen et al. (2020) findings, which indicated educational-purpose applications are better tailored to learners' needs. A plausible explanation might be that, applications for general purposes used in current study are instant messaging tools, e.g., WeChat, WhatsApp and telegram, that featured in peer interactions and collaborations (Li et al., 2019; Li et al., 2021). Reading supported by the interactive and collaborative MALL applications would sustain EFL learners' attention and motivate their reading interest and engagement (Chen & Lin, 2016).

Instructional approaches. A scrutiny of the beneficial effects indicated that all the five instructional approaches are effective, and drill and practice (g = 1.079) and multimedia learning (g = 0.869) obtained high effects, while

game-based learning (g = 0.786), personalized learning (g = 0.785) and collaborative learning (g = 0.601) had moderate effects. This might be attributed to the overwhelming influence of the traditional "behaviorist, teachercentred, transmission model of instruction" (Burston, 2014, p. 344) that normally used MALL for multimedia glosses (Chen et al., 2011; Yanagisawa et al., 2020). In other words, MALL technologies were often used informally as a multimedia complement for formal classroom instruction, highlighting the drill and practice features to develop FL learners' reading ability (García Botero et al., 2019). Aside from drill and practice and multimedia learning, other effective approaches should receive equal attention, since game-based learning, personalized learning and collaborative learning also obtained significantly moderate beneficial effects.

Measured outcome types. Measured outcome types consist of standardized tests and researcher-designed tests. Although effect size of researcher-designed tests (ES = 0.868) was slightly higher than that of standardized tests (ES = 0.734), no significant difference existed between the two measured outcome types, corroborating previous findings that compared the moderating effect between standardized tests and researcher-designed tests in language learning anxiety (Li, 2021a) and CALL for writing quality (Xu et al., 2019). This result suggests that both standardized tests and researcher-designed tests could warrant a good reliability to measure EFL learners' reading performance.

6. Implications

Some practical implications for teachers, providers, designers and researchers are inferred from the major findings that follow.

6.1. Implications for teachers/providers

For teachers or provides, reading materials should be adjusted to EFL learners' current proficiency levels. Teachers or providers should take learners' proficiency levels into consideration, before distributing reading tasks or assignments to learners. As low proficiency learners often accompanied with small vocabulary size and limited morphosyntactic knowledge (Droop & Verhoeven, 2003), a simplified version with high frequency words and easy-to-understand grammatical structures would be preferred. Besides, reading materials should be tailored to learners' cognitive development. Our findings showed that adult learners performed better than the children and adolescent learners when using MALL for L2 reading comprehension, suggesting that teachers and/or MALL providers should also consider learners' cognitive development when preparing the reading materials. For young readers, multimedia reading texts that integrated audio, pictorial and textual materials altogether would be better than unimodal text-only materials (Li, 2021b; Mayer, 2009). Importantly, given that all five instructional approaches were found to be significantly effective but only differed in the magnitude of effect sizes, MALL applications should be integrated into curriculums with tailored approaches depending on the educational need and purpose (Li, 2022).

6.2. Implications for designers

MALL application designers should consider learners' personalized needs and develop the easy-to-use and userfriendly interface of MALL applications to "automatically adjust text display type in different reading contexts to promote reading comprehension, sustained attention, or reduce cognitive load based on reading contexts" (Chen & Lin, 2016, p. 568; Gutiérrez-Colón et al., 2020). Moreover, since the general-purpose applications featured in interactions and collaborations have a larger effect size, designers should also consider the integration of interactive and collaborative features into learning materials for the application designs (Li, 2022; Li et al., 2021). As the beneficial effects of larger screens would be higher than the smaller ones, designers should also consider using "appropriate text display type for mobile reading in different contexts, adjusting it to the reading context to improve reading comprehension, attention, or cognitive load" (Gutiérrez-Colón et al., 2020, p. 7).

6.3. Implications for researchers

Given that MALL devices are "ideal tools for creating an interactive, collaborative and ubiquitous environment for language learning" (Chen, 2013, p. 20), researchers should try to establish informal outside-of-the-classroom learning settings for EFL learners, so that they can plan, monitor, manage and autonomously self-regulate their learning process. Furthermore, researchers should also consider adopting longer intervention durations to

improve the reliability and ecological validity of research design. With short-term durations (e.g., one session or less than 1 week), it would be rather difficult to ensure whether the effects "are produced by the features of MALL devices rather than by the experience of technology novelty" (Sung et al., 2016, p. 265-266). Moreover, as far as the short-term durations are concerned, the integrative effect between MALL devices and the curriculum would also not be satisfactorily achieved.

7. Conclusion

The results showed that the overall effect size was significantly large, suggesting the use of MALL for L2 reading comprehension is more effective than traditional methods. For moderator analysis results, the intervention settings and intervention durations were found to be significant moderators, while others did not find a significant moderating effect.

There are some limitations to be addressed though. On the one hand, due to the strict inclusion/exclusion criteria, some of the highly related empirical studies that did not report sufficient statistical information for effect size calculation, unpublished studies and publications written in other languages were not included, resulting in only 20 independent studies that met the inclusion criteria, which might affect the comprehensiveness of the meta-analysis results. Future study should include more eligible empirical studies with more keyword combinations and wider time range. On the other hand, the moderators included in this study were based on several existing meta-analyses of MALL for language learning (Chen et al., 2020; Sung et al., 2015; Sung et al., 2016), other potential moderators were not considered in the literature. Thus, researchers should include more potential moderators with sufficient information for calculation in the future.

Acknowledgement

This research was supported by the Social Science Foundation of Hunan Province (grant number 20ZDB005). The author would like to thank the anonymous reviewers and Prof. Yu-Fen Yang for their constructive and insightful comments.

References

Articles marked with an asterisk (*) were analyzed in the meta-analysis.

Alemi, M., & Lari, Z. (2012). SMS vocabulary learning: A Tool to promote reading comprehension in L2. *International Journal of Linguistics*, 4(4), 275–287. https://doi.org/10.5296/ijl.v4i4.2318

*Ataee, E., Fatemi, M. A., & Ashraf, H. (2015). The Effect of short message service on Iranian lower intermediate EFL learners' reading comprehension through skimming and guessing strategies awareness. *Theory and Practice in Language Studies*, 5(7), 1428–1437. https://doi.org/10.17507/tpls.0507.15

Biernacki, P., & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referral sampling. *Sociological Methods & Research*, *10*(2), 141–163. https://doi.org/10.1177/004912418101000205

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2005). *Comprehensive meta-analysis version 2.0* (Computer software). Biostat.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Introduction to meta-analysis. John Wiley & Sons, Ltd.

Burston, J. (2014). MALL: The Pedagogical challenges. *Computer Assisted Language Learning*, 27(4), 344–357. https://doi.org/10.1080/09588221.2014.914539

Burston, J. (2015). Twenty years of MALL project implementation: A Meta-analysis of learning outcomes. *ReCALL*, 27(1), 4–20. https://doi.org/10.1017/S0958344014000159

Chen, Z., Chen, W., Jia, J., & An, H. (2020). The Effects of using mobile devices on language learning: A Meta-analysis. *Educational Technology Research and Development*, 68, 1769–1789. https://doi.org/10.1007/s11423-020-09801-5

^{*}Chen, C., & Hsu, S. (2008). Personalized intelligent mobile learning system for supporting effective English learning. *Educational Technology & Society*, *11*(3), 153–180.

Chen, C. M., & Lin, Y. J. (2016). Effects of different text display types on reading comprehension, sustained attention and cognitive load in mobile reading contexts. *Interactive Learning Environments*, 24(3), 553–571. https://doi.org/10.1080/10494820.2014.891526

*Chen, N., Teng, D. C., Lee, C., & Kinshuk. (2011). Augmenting paper-based reading activity with direct access to digital materials and scaffolded questioning. *Computers & Education*, 57, 1705–1715. https://doi.org/10.1016/j.compedu.2011.03.013

Chen, X. B. (2013). Tablets for informal language learning: Student usage and attitudes. *Language Learning & Technology*, 17(1), 20–36. https://scholarspace.manoa.hawaii.edu/bitstream/10125/24503/1/Chen%20X.pdf

Cho, K., Lee, S., Joo, M., & Becker, B. (2018). The Effects of using mobile devices on student achievement in language learning: A Meta-analysis. *Education Sciences*, 8(3), 1–16. https://doi.org/10.3390/educsci8030105

Davis, J. N., & Lyman-Hager, M. A. (1997). Computers and L2 reading: Student performance, student attitudes. *Foreign Language Annals*, 30(1), 58–72. https://doi.org/10.1111/j.1944-9720.1997.tb01317.x

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second-language learners. *Reading Research Quarterly*, *38*, 78–103. https://doi.org/10.1598/RRQ.38.1.4

Duman, G., Orhon, G., & Gedik, N. (2014). Research trends in mobile assisted language learning from 2000 to 2012. *ReCALL*, 27(2), 197–216. https://doi.org/10.1017/S0958344014000287

García Botero, G., Questier, F., & Zhu, C. (2019). Self-directed language learning in a mobile-assisted, out-of-class context: Do students walk the talk? *Computer Assisted Language Learning*, 32(1-2), 71–97. https://doi.org/10.1080/09588221.2018.1485707

*Gheytasi, M., Azizifar, A., & Gowhary, H. (2015). The Effect of smartphone on the reading comprehension proficiency of Iranian EFL learners. *Procedia - Social and Behavioral Sciences*, 199, 225–230. https://doi.org/10.1016/j.sbspro.2015.07.510

Grami, S., & Hashemian, M. (2017). Effect of paper and e-dictionaries on Iranian EFL learners' reading comprehension. *International Journal of Applied Linguistics & English Literature*, 6(5), 199–205. https://doi.org/10.7575/aiac.ijalel.v.6n.5p.199

Gutiérrez-Colón, M., Frumuselu, A. D., & Curell, H. (2020). Mobile-assisted Language learning to enhance L2 reading comprehension: A Selection of implementation studies between 2012–2017. *Interactive Learning Environments*, forthcoming. https://doi.org/10.1080/10494820.2020.1813179

Hazaea, A. N., & Alzubi, A. A. (2016). The Effectiveness of using mobile on EFL learners' reading practices in Najran University. *English Language Teaching*, 9(5), 8–21. https://doi.org/10.5539/elt.v9n5p8

*Hsu, C., Hwang, G., & Chang, C. (2013). A Personalized recommendation-based mobile learning approach to improving the reading performance of EFL students. *Computers & Education*, 63, 327–336. https://doi.org/10.1016/j.compedu.2012.12.004

Hwang, G., & Fu, Q. (2019). Trends in the research design and application of mobile language learning: A Review of 2007-2016 publications in selected SSCI journals. *Interactive Learning Environments*, 27(4), 567–581. https://doi.org/10.1016/j.compedu.2018.01.004

Keezhatta, M. S., & Omar, A. (2019). Enhancing reading skills for Saudi secondary school students through mobile assisted language learning (MALL): An Experimental study. *International Journal of English Linguistics*, 9(1), 437–447. https://doi.org/10.5539/ijel.v9n1p437

*Khubyari, L., & Narafshan, M. H. (2016). A Study on the impact of MALL (mobile assisted language learning) on EFL learners' reading comprehension. *International Journal of English Language Teaching*, 4(2), 58–69.

Klimova, B., & Zamborova, K. (2020). Use of mobile applications in developing reading comprehension in second language acquisition: A review study. *Education Sciences*, *10*(12), 1–11. https://doi.org/10.3390/educsci10120391

^{*}Lan, Y., Sung, Y., & Chang, K. (2013). From particular to popular: Facilitating EFL mobile-supported cooperative reading. *Language Learning & Technology*, *17*(3), 23–38.

^{*}Lan, Y., Sung, Y., & Chang, K. (2009). Let us read together: Development and evaluation of a computer-assisted reciprocal early English reading system. *Computers & Education*, 53(4), 1188–1198. https://doi.org/10.1016/j.compedu.2009.06.002

Li, R. (2021a). Foreign language reading anxiety and its correlates: A Meta-analysis. *Reading and Writing*, forthcoming. https://doi.org/10.1007/s11145-021-10213-x

Li, R. (2021b). Does game-based vocabulary learning APP influence Chinese EFL learners' vocabulary achievement, motivation and self-confidence? Sage Open, 11(1), 1–12. https://doi.org/10.1177/21582440211003092

Li, R. (2022). Research trends of blended language learning: A Bibliometric synthesis of SSCI-indexed journal articles during 2000–2019. *ReCALL*, forthcoming. https://doi.org/10.1017/S0958344021000343

Li, R., Meng, Z., Tian, M., Zhang, Z., & Xiao, W. (2021). Modelling Chinese EFL learners' flow experiences in digital game-based vocabulary learning: The roles of learner and contextual factors. *Computer Assisted Language Learning*, *34*(4), 483–505. https://doi.org/10.1080/09588221.2019.1619585

Li, R., Meng, Z., Tian, M., Zhang, Z., Ni, C., & Xiao, W. (2019). Examining EFL learners' individual antecedents on the adoption of automated writing evaluation in China. *Computer Assisted Language Learning*, *32*(7), 784–804. https://doi.org/10.1080/09588221.2018.1540433

Li, Y., & Hafner, C. A. (2022). Mobile-assisted vocabulary learning: Investigating receptive and productive vocabulary knowledge of Chinese EFL learners. *ReCALL*, 34(1), 66–80. https://doi.org/10.1017/S0958344021000161

^{*}Lin, C. (2014). Learning English reading in a mobile-assisted extensive reading program. *Computers & Education*, 78, 48–59. https://doi.org/10.1016/j.compedu.2014.05.004

*Lin, C. (2017). Learning English with electronic textbooks on tablet PCs. *Interactive Learning Environments*, 25(8), 1035–1047. https://doi.org/10.1080/10494820.2016.1242505

Lin, C., Lin, V., Liu, G., Kou, X., Kulikova, A., & Lin, W. (2020). Mobile-assisted reading development: A Review from the Activity Theory perspective. *Computer Assisted Language Learning*, 33(8), 833–864. https://doi.org/10.1080/09588221.2019.1594919

Lin, J., & Lin, H. (2019). Mobile-assisted ESL/EFL vocabulary learning: A Systematic review and meta-analysis. *Computer Assisted Language Learning*, 32(8), 878–919. https://doi.org/10.1080/09588221.2018.1541359

Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. SAGE.

Mayer, R. E. (2009). Multimedia learning (2nd ed.). Cambridge University Press.

^{*}Mays, B. R., Yeh, H., & Chen, N. (2020). The Effects of using audience response systems incorporating student-generated questions on EFL students' reading comprehension. *The Asia-Pacific Education Researcher*, 29(6), 553–566. https://doi.org/10.1007/s40299-020-00506-0

Moon, A. L., Francom, G. M., & Wold, C. M. (2021). Learning from versus learning with technology: Supporting constructionist reading comprehension learning with iPad applications. *TechTrends*, 65(1), 79–89. https://doi.org/10.1007/s11528-020-00532-1

Motallebzadeh, K., & Ganjali, R. (2011). SMS: Tool for L2 vocabulary retention and reading comprehension ability. *Journal of Language Teaching and Research*, 2(5), 1111–1115. https://doi.org/10.4304/jltr.2.5.1111-1115

*Naderi, S., & Akrami, A. (2018). EFL learners' reading comprehension development through MALL: Telegram groups in focus. *International Journal of Instruction*, *11*(2), 339–350. https://doi.org/10.12973/iji.2018.11223a

Plonsky, L. (2011). The effectiveness of second language strategy instruction: A Meta-analysis. *Language Learning*, 61(4), 993–1038. https://doi.org/10.1111/j.1467-9922.2011.00663.x

Plonsky, L., & Oswald, F. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, 64, 878–912. https://doi.org/10.1111/lang.12079

*Priyanti, N. W. I., Santosa, M. H., & Dewi, K. S. (2019). Effect of quizizz towards the eleventh-grade English students' reading comprehension in mobile learning context. *Language and Education Journal Undiksha*, 2(2), 71–80. https://doi.org/10.23887/leju.v2i2.20323

Reiber-Kuijpers, M., Kral, M., & Meijer, P. (2021). Digital reading in a second or foreign language: A Systematic literature review. *Computers & Education*, *163*, 1–26. https://doi.org/10.1016/j.compedu.2020.104115

Rosenthal, R. (1991). Meta-analytic procedures for social research. Sage.

Shadiev, R., Liu, T., & Hwang, W. Y. (2020). Review of research on mobile-assisted language learning in familiar, authentic environments. *British Journal of Educational Technology*, 51(3), 709–720. https://doi.org/10.1111/bjet.12839

Singhal, M. (1998). A Comparison of L1 and L2 reading: Cultural differences and schema. *The Internet TESL Journal*, 4(10). http://iteslj.org/Articles/Singhal-ReadingL1L2.html

*Sofiana, N., & Mubarok, H. (2020). The Impact of English game-based mobile application on students' reading achievement and learning motivation. *International Journal of Instruction*, *13*(3), 247–258. https://doi.org/10.29333/iji.2020.13317a

Sung, Y., Chang, K., & Yang, J. (2015). How effective are mobile devices for language learning? A Meta-analysis. *Educational Research Review*, *16*, 68–84. https://doi.org/10.1016/j.edurev.2015.09.001

Sung, Y., Chang, K., & Liu, T. (2016). The Effects of integrating mobile devices with teaching and learning on students' learning performance: A Meta-analysis and research synthesis. *Computers & Education*, 94, 252–275. https://doi.org/10.1016/j.compedu.2015.11.008

^{*}Wang, Y. (2017). Integrating self-paced mobile learning into language instruction: Impact on reading comprehension and learner satisfaction. *Interactive Learning Environments*, *25*(3), 397–411. https://doi.org/10.1080/10494820.2015.1131170

Wang, S., & Higgins, M. (2005). Limitations of mobile phone learning. Paper presented at the *IEEE International Workshop* on Wireless and Mobile Technologies in Education (WMTE'05), Tokushima, Japan. https://doi.org/10.1109/WMTE.2005.43

Whitford, V., & Joanisse, M. F. (2018). Do eye movements reveal differences between monolingual and bilingual children's first-language and second-language reading? A Focus on word frequency effects. *Journal of Experimental Child Psychology*, *173*, 318–337. https://doi.org/10.1016/j.jecp.2018.03.014

Wilson, D. B. (2019). Systematic coding for research synthesis. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-Analysis (3rd edition)* (pp.153–172). Russell Sage Foundation.

^{*}Wu, T., Sung, T., Huang, Y., & Yang, C. (2010). Location awareness mobile situated English reading learning system. *Journal of Internet Technology*, *11*(7), 923–933. https://doi.org/10.6138/JIT.2010.11.7.05

^{*}Wu, T., Sung, T., Huang, Y., Yang, C., & Yang., J. (2011). Ubiquitous English learning system with dynamic personalized guidance of learning portfolio. *Educational Technology & Society*, *14*(4), 164–180. https://drive.google.com/file/d/1dpSCn0MNEUUSh9DhgdwhJm99JITw_DBb/view

Xu, Z., Banerjee, M., Ramirez, G., Zhu, G., & Wijekumar, K. (2019). The Effectiveness of educational technology applications on adult English language learners' writing quality: A Meta-analysis. *Computer Assisted Language Learning*, 32(1-2), 132–162. https://doi.org/10.1080/09588221.2018.1501069

Yanagisawa, A., Webb, S., & Uchihara, T. (2020). How do different forms of glossing contribute to L2 vocabulary learning from reading? A Meta-regression analysis. *Studies in Second Language Acquisition*, 42(2), 411–438. https://doi.org/10.1017/S0272263119000688

Engineering Students' Readiness for Online Learning Amidst the COVID-19 Pandemic: Scale Validation and Lessons Learned from a Developing Country

Md Abdullah Al Mamun^{1*}, Md Akbar Hossain², Sayedus Salehin³, Md. Shahadat Hossain Khan¹ and Mahbub Hasan¹

¹Department of Technical and Vocational Education, Islamic University of Technology, Bangladesh // ²School of Business and Digital Technologies, Manukau Institute of Technology, New Zealand // ³Department of Mechanical and Production Engineering, Islamic University of Technology, Bangladesh // a.mamun@iutdhaka.edu // akbar.hossain@manukau.ac.nz // sayedus@iut-dhaka.edu // shkhants@iut-dhaka.edu // m.hasan@iut-dhaka.edu

*Corresponding author

(Submitted June 10, 2021; Revised September 19, 2021; Accepted November 4, 2021)

ABSTRACT: The recent outbreak of the COVID-19 pandemic forced education institutes to shift to an internetbased online delivery mode. This unique situation accelerates a long-standing issue of digital inequality among the students in education and warrants a concentrated study to investigate students' readiness for learning in online environment. This study developed an instrument to meticulously measure the students' readiness for online learning in a pandemic situation. The proposed model consists of (a) motivation, (b) self-efficacy, and (c) situational factors. The proposed model was validated with the engineering students (for pilot study N = 68 and main study N = 988) from several universities in Bangladesh. To validate the underlying relationships between the latent constructs, an exploratory factor analysis (EFA) was performed followed by structural equation modelling (SEM) for the construct validity of the measurement model and to assess the model fit. The findings showed that besides motivation and self-efficacy, the situational factors describing the contextual dynamics emerging from the COVID-19 significantly influenced the student's online readiness. We argue that digital inequality is an important factor influencing student readiness for online learning.

Keywords: Online learning readiness, COVID-19 pandemic, Bangladesh, Engineering education, Structural Equation Modelling, Situational factors, Digital inequality

1. Introduction

Bangladesh, being a high risk and country vulnerable to the COVID-19 pandemic (Hossain et al., 2020; Monjur & Hassan, 2020), took several measures to combat transmission of the virus. The most immediate measure introduced by the country was to regulate the practice of "social distancing" (Yeasmin et al., 2020) to flatten the curve of COVID-19 transmission. As a result, all educational institutions were closed across the country. Social distancing became the "new normal" for students and the usual comradeship of campus life disappeared. This has drastically impacted on Bangladesh's educational system, resulting in a loss of learning opportunities. Roughly 3.7 million students and a million teachers in the higher education sector are reportedly now stuck at home (Ahmed, 2020).

To minimize interruption caused by the COVID-19 pandemic, engineering universities in Bangladesh acted quickly to shift all face-to-face lectures to a home-based online distance learning mode using learning platforms such as Google classroom, Moodle, Zoom etc. Some universities even consider adopting flipped learning approach because of its effectiveness compared to traditional instructions reported in the recent literature (Chang et al., 2020; Galindo-Dominguez, 2021; Zheng et al., 2020). This paradigm shift from face-to-face learning to online distance mode creates two major complexities. Firstly, academic matters such as delivery, teachers' expertise, student preparedness, and engagement within this new virtual learning space must all be addressed (Ioannou & Ioannou, 2020; Khtere & Yousef, 2021). The second issue, perhaps more sensitive, relates to the physical and psychological wellbeing of the students. The absence of social and physical interaction has adverse effects on students' wellbeing (Twenge et al., 2019). Nevertheless, engineering universities are continuing to shift course delivery to fully-fledged online learning environments as no other viable solutions are available. Students get little time to cope with this "new normal" in their educational lives.

Therefore, an important question requires immediate attention: To what extent are the engineering students of Bangladesh ready for the online classes that are replacing face-to-face learning during the COVID-19 pandemic? Because the event is unique, research into understanding student readiness for online learning in a pandemic

situation is only starting to emerge, and no reported research has been found in the context of engineering education in Bangladesh. Though several studies attempted to measure students' readiness for online learning (Arthur-Nyarko et al., 2020; Yu, 2018), none of them fully address the factors relating to an emergency. Chung et al. (2020) measured students' online learning readiness amidst the COVID-19 pandemic, nonetheless, they did not address the situational and context specific factors that emerged due to the pandemic. Thus, a careful understanding of the current pandemic situation and a reconceptualisation of the dimensions and constructs of the students' readiness for online learning is warranted.

For this reason, the current study develops and validates a more specific instrument that can be used to measure the students' readiness for online learning in a pandemic situation. Secondly, this study investigates how demographic factors influence the online learning readiness of engineering students of Bangladesh during the pandemic. Thus, this study sought to answer the following two questions in the context of the current pandemic caused by COVID-19:

- RQ1: What is the reliability, validity, and model fit evidence of the survey scale to assess engineering students' readiness for online learning?
- RQ2: To what extent are engineering students of Bangladesh (in terms of gender, level of study, place of living, and university type) ready to learn in online environments?

2. Reconceptualising the constructs of students' readiness in the pandemic situation

2.1. Motivation and self-efficacy: Two key constructs of students' online readiness

In previous literature, motivation was identified as the most crucial construct of students' readiness for online learning (Chung et al., 2020; Xiong et al., 2015; Yu, 2018). In the current pandemic situation, this has similarly become the primary factor for students to engage successfully in remote learning. The absence of social structure, close interactions, easy access to teachers and peers in online learning during COVID-19 pandemic may influence students' motivation and readiness to learn in this manner (Allam et al., 2020).

Motivation, as conceptualised in our study, delineates students' willingness to use online learning platforms during the COVID-19 pandemic. Guided by self-determination theory (Ryan & Deci, 2000), we considered students' intrinsic motivation i.e., interest or enjoyment, and extrinsic motivation i.e., perceived usefulness and reinforcement, to be the key aspects to evaluate students' motivation in our study. Self-determination theory further contends that students' connectedness with their teachers and peers are a vital component of student motivation. Previous literature also demonstrates the importance of engaging in human-human interactions and the sense of being part of a learning community for effective learning in online settings (Joksimović et al., 2015). Students get a feeling of connectedness to other students through online learning communities, and this contributes to meaningful learning experiences (Cho & Tobias, 2016).

COVID-19 also requires students to heavily depend on technology and to equip themselves with computer/internet literacy for successful online participation (Allam et al., 2020). COVID-19 entails students to have self-efficacy i.e., knowledge of and competencies in using modern technologies to achieve the educational objectives determined by their academic institutions (Lai, 2011). Even before the pre-COVID era, self-efficacy is considered as an important skill for learning in contemporary online settings (Hung et al., 2010). Early literature refers to self-efficacy is considered as a major driving factor in preparing students for online learning (Hung et al., 2010; Xiong et al., 2015) and that social and technical competency, two key dimensions of self-efficacy for student learning, are highly associated with online readiness and satisfaction (Yu, 2018; Yu & Richardson, 2015).

2.2. Situational factors: The emerging constructs for students' online readiness

Miglani and Awadhiya (2017) pointed out that the availability of digital resources and the ability to use and benefit from these are the key factors that characterize digital inequality. Based on the notion of digital inequality accelerated by the COVID-19 pandemic, several key dimensions with increased relevancy to students' readiness for online learning become apparent. In this study, we identified these dimensions under a common construct named "situational factors."

The first factor we conceptualise is the availability and access to the digital resources amidst a pandemic situation. Research shows that low-income families are suffering the most from the COVID-19 economic crisis because they have fewer and lower quality digital appliances (Fernandes, 2020). Bangladesh is not an exception here. Due to their low socio-economic status, many students in Bangladesh do not have the modern devices to readily adjust to the technology based "new normal" life. Instead, research shows that use of outdated devices, as is the supposed case for the majority students of Bangladesh, results in delays in connecting to online resources and an overall less satisfying experience (Beaunoyer et al., 2020). Also, the increased cost of internet data and poor connectivity remains a serious threat for technology adoption in Bangladesh (Ullah et al., 2021). As a result, students get fewer opportunities to access, engage with, and experience modern technologies.

Second is the "learning atmosphere" in the home environment - a unique and unprecedented context emerging because of COVID-19 lockdown. Neuwirth et al. (2020) reasoned that some issues are exacerbated by underlying conditions of disparity of available resources triggered by the COVID-19 pandemic. These include the lack of a calm and peaceful study space within the home environment which can help students to learn in comfort and with privacy. However, a positive learning atmosphere is not simply silence: it is a complex-to-describe combination of sense experience and feelings shaped by underlying spatial organization, structures, social rules, and interactions governed by the environment (Cox, 2017). Often too, close proximities with other family members trigger disturbances, and students can be reluctant to use a webcam during classes which may expose their socioeconomic and living conditions (Neuwirth et al., 2020).

Leaning atmosphere at home are important sources for the development of positive self-efficacy which regulate students' learning in online environment (Bonanati & Buhl, 2021). Research shows that factors within the home learning atmosphere can predict students' self-efficacy (Bonanati & Buhl, 2021; Rohatgi et al., 2016). In contrast, when student experiences poor learning environments it affects their self-efficacy development and learning outcomes (Khine et al., 2020). For example, many students are facing difficulties in online assignment submission and tasks accomplishment during pandemic because of poor learning atmosphere (Bisht et al., 2020). In brief, the learning atmosphere is a crucial ingredient to stimulate student motivation (Pamungkas, 2019). Evidence indicates that a supportive learning atmosphere has a major influence on student self-efficacy and attitudes toward learning (Han & Ellis, 2021; Kokoç et al., 2021).

Third is the institutional support which can reduce the huge academic gap emerged due to remote learning. In fact, institutional support and quality education are linked in a significant way (Ullah et al., 2021). Educational institutions should facilitate student learning by providing emotional support and necessary information to help alleviate common challenges faced by online learners (Huang et al., 2020). The home confinement triggered by COVID-19 limits access to the faster networks readily available at educational institutions (Beaunoyer et al., 2020). When students are deprived of such facilities, educational institutions should subsidise the internet cost for students from low-income families. These types of supports can significantly help students to prepare themselves for online learning.

In contrast, the poor institutional support services may intensify these problems and affect student self-efficacy (Richardson et al., 2021). Irani et al. (2014) even claimed that institutions should consider multiple ways to support online students to mitigate the feeling of loneliness and separation from their peers and teachers. These supports can help online learners navigate important administrative, technical, financial, and other educational challenges while also increase students' self-efficacy and improve student retention in online courses in the long run (Trespalacios et al., 2021). All these situational specific factors therefore signify the importance of students' preparedness, motivation and their self-efficacy for online learning and their continuous intentions to use (Wang & Lin, 2021).




Based on the understanding of different constructs of students' readiness amidst a pandemic situation, we therefore propose a reconceptualised model of students' readiness for online learning (Figure 1). This model consists of three key components: motivation, self-efficacy, and situational factors. Further, in this model we conceptualise situational factors as a combination of three sub-constructs: digital access, learning atmosphere and institutional support.

3. Research methods

3.1. Scale development

The scale development process was finalised in four different phases suggested by DeVillis (2016). First, we generated items based on related previous research employing a five-point Likert scale. Second, we modified and refined the items based on experts' feedback. Third, we conducted a pilot study with a sample of 68 students to check initial internal consistency and inter-item correlations of the items. Finally, we tested the reliability and validity of the survey scale using a larger student sample in the actual study.

Previous studies showed positive correlations between different motivational factors originated from self-determination theory such as interest, perceived usefulness, reinforcement, connectedness, and students' level of online readiness (Hung et al., 2010; Xiong et al., 2015). Therefore, we adapted seven items from Hung et al. (2010), Xiong et al. (2015) and *Intrinsic Motivation Inventory (IMI)* rooted in the self-determination theory (Ryan & Deci, 2000) to measure student interest; seven items from *IMI* and Xiong et al. (2015) to measure perceived usefulness; four items (two from Xiong et al. (2015) and two newly created) to measure reinforcement. Finally, we adapted eight items from *IMI* to measure students' relatedness in online learning.

To measure students' self-efficacy, we adapted four items from Yu and Richardson (2015) to measure students' technical competency and ten items from Hung et al. (2010) to measure social competency. The final constructs in our study are characterized as *situational factors* which describe the contextual dynamics emerging from the COVID-19 pandemic. We conceptualised this construct as the combination of three sub-constructs i.e., digital access, learning atmosphere and institutional support. Thus, we have created twelve new items under the situational factors (four items for learning atmosphere, four items for institutional support and four items for digital access). In total, there were 52 items in the initial survey instrument (see Appendix).

3.2. Research contexts and participants

The researchers started distributing the online survey during the peak of COVID-19 at the beginning of June 2020, when all the higher educational institutes of Bangladesh had already started online teaching. The survey was administered nationwide in a total of 23 universities. To achieve a representative sample for the study, participants were invited from all three types of universities: public (government funded), private, and international (funded by international donor agencies).

3.3. Data collection and preparation

Initially a total of 1038 responses were collected using Google form. After a rigorous data screening process, 988 responses were found to be valid. The data set had been scrutinized for missing values, normality, and outliers. The summary of the participants' demographic data can be found in the supplementary dataset at the end of the document. The reliability and descriptive statistics of the data set are shown below (Table 1).

10	<i>uble 1.</i> Reliability and descriptive	he descriptive statistics of the theoretical constructs					
Constructs		Mean	Std. Deviation	Skewness	Kurtosis		
Motivation	Interest	18.45	7.32	.319	757		
$\alpha = 0.964$	Usefulness	18.59	7.56	.311	856		
	Reinforcement	11.58	4.18	.076	824		
	Connectedness	21.67	6.94	.176	584		
Self-efficacy	Technology competency	13.93	3.99	403	440		
$\alpha = 0.926$	Social competency	29.68	9.21	.101	608		
Situational Factors	Learning atmosphere	12.81	4.06	043	769		
$\alpha = 0.868$	Institutional support	13.18	4.16	265	654		
	Digital access	12.46	3.93	.038	610		

Table 1. Reliability and descriptive statistics of the theoretical constructs

Table 1 shows that the coefficient alpha values were well above 0.8 which showed very good internal consistency among the items (Blunch, 2008). Our data set also met assumptions of multivariate normality as both skewness (< 3.0) and kurtosis (< 10) are within the range (Kline, 2016).

3.4. Data analysis

To answer the RQ1, we first conducted an exploratory factor analysis (EFA) to determine the relationships between latent variables reflected in the items of the survey instrument (Hair et al., 2010). Table 2 shows the recommended index values for EFA analysis used in this study.

Tuble 2. KG	commended mack values for Er	A used in this study
Indicators	Recommended value	Source
Kaiser-Meyer-Olkin (KMO)	> 0.70	Hutcheson and Sofroniou (1999)
Bartlett's test of sphericity	Significant at $p < 0.001$	Field (2013)
Satisfactory communalities values	> 0.50	Field (2013)
Total variance explained	> 50%	Podsakoff and Organ (1986)
The variance for the first factor	< 50%	Podsakoff and Organ (1986)
Factor loading for items	> 0.50	Hair et al. (2016)

Table 2. Recommended index values for EFA used in this study

Second, we conducted confirmatory analysis (CFA) to examine the reliability, convergent validity, and discriminant validity of our proposed model. Third, in the structural model, we assessed the model fit against several tests and fit indices recommended by literature (see Table 6 for details). Finally, to address RQ2, we ran a multivariate analysis of variance (MANOVA) to explore the students' readiness for online learning with regards to different demographic variables.

4. Results

4.1. Exploratory factor analysis

We used Monte Carlo software program for parallel analysis to identify the exact number of components to best reflect the underlying relationship among the variables. We kept only those components with the eigenvalues greater than the randomly generated data from parallel analysis (see supplementary dataset). For a cleaner solution, items with high communalities and factor loadings (greater than 0.5) were retained in EFA. In this process, a total of 39 survey items were retained for the EFA model. EFA suggested a four-factors model comprising motivation, self-efficacy, learning atmosphere and institutional support.

Tuble 5. men	Tactor correla		ind renability		Tillouel
Factors	1	2	3	4	Reliability (Cronbach α)
1. Motivation	1.000				0.971
2. Self-efficacy	.461	1.000			0.863
3. Learning atmosphere	.633	.488	1.000		0.860
4. Institutional support	.580	.333	.581	1.000	0.853
Sampling Adequacy					
КМО		0.9	980		
Bartlett's tests of sphericity	0.000^{***}				
Total Variance Explained	62.62%				

Table 3. Inter factor correlation matrix and reliability of the EFA model[#]

Note. [#]Extraction Method: Principal Component Analysis; Rotation Method: Promax with Kaiser Normalization. ^{***}Significant at p < .001.

Table 3 shows excellent internal consistency (Cronbach's alpha) of the items in the four factors EFA model. Discriminant validity is also ensured as no cross loading of the items are observed in more than one factor and inter factor correlations are below 0.70. Bartlett's tests of sphericity were found to be significant (0.000; p < .001) with excellent KMO value (.980), suggesting the suitability of factor analyses.

4.2. Measurement model

In validating the measurement model with confirmatory factor analysis (CFA), we found some problematic items and therefore, following suggested data-analysis practices (MacCallum et al., 1996), we retain 30 items for the final model. Table 4 shows excellent composite reliability, high factor loadings and standardized regression weights (greater than .05 at p < .001) which support the convergent validity of the model (Hair et al., 2010). The average variance extracted (AVE) (greater than .50) also confirm the convergent validity (Fornell & Larcker, 1981).

Table 4.	Convergent ar	nd discrimin	nant validity	y of the measu	rement mode	el	
Constructs	CR	AVE	MSV	1	2	3	4
1. Motivation	0.967	0.623	0.442	0.790			
2. Self-efficacy	0.824	0.540	0.528	0.665^{***}	0.735		
3. Learning atmosphere	0.831	0.555	0.528	0.659^{***}	0.727^{***}	0.745	
4. Institutional Support	0.856	0.600	0.425	0.652^{***}	0.607^{***}	0.533***	0.775

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

The correlations of the constructs and the square root of the AVE on the diagonal (in bold numbers) are shown in Table 4. As revealed, all square root of AVEs is greater than the inter factor correlations and all AVEs are greater than the MSVs (maximum shared variance) (Fornell & Larcker, 1981). Thus, our model met the criteria of discriminant validity. Further, the heterotrait-monotrait (HTMT) ratio of correlations in Table 5 are below .850 showing a strict discriminant validity between the factors (Fornell & Larcker, 1981; Henseler et al., 2015).

	Table 5. HTMT	Analysis		
Constructs	1	2	3	4
1. Motivation				
2. Self-efficacy	0.687			
3. Learning atmosphere	0.671	0.757		
4. Institutional support	0.668	0.628	0.562	

In sum, the evaluation of the measurement model suggested that all items are reliable and met the conditions of convergent and discriminant validity.

4.3. Structural model

SRMR

Hu and Bentler (1999) state that a RMSEA value less than 0.07, and CFI and TLI values greater than 0.90 indicate good fit of a model. In our study, the value of the RMSEA coefficient is 0.063, and other indicators (CFI, TLI, IFI, and NFI) are all above 0.90 which indicate a good fit for the model. SRMR fit index is also smaller than 0.10, further confirming. Thus, we conclude that our model met all the recommended levels of fit indices (Table 6).

Table 6. Recommended values of the fit indices and the corresponding results of the proposed model						
Fit Index	Admissibility	Source	Result	Fit		
CMIN/DF	< 5.0	Hu and Bentler (1999); Kline (2016)	(1954.32/399)	Yes		
			= 4.898			
RMSEA	< 0.08	Hu and Bentler (1999)	0.063	Yes		
CFI	> 0.90	Hu and Bentler (1999)	0.929	Yes		
TLI	> 0.90	Hu and Bentler (1999);	0.923	Yes		
IFI	> 0.90	Hu and Bentler (1999);	0.929	Yes		
NFI	> 0.80	Bentler and Bonett (1980); Schumacker and Lomax	0.912	Yes		

We also assessed for multicollinearity issue using variance inflation factor (VIF) and found that all the values are between 1.903 and 3.550. Thus, the VIF values met the criteria to support the structural model (Hair et al., 2016; Kline, 2016).

0.045

Yes

(2010)

Hu and Bentler (1999)

< 0.10

Table 7 shows the path coefficients and path significances revealing that all values are significant between the factors (at p < .001).

	Table 7. Model path analysis							
Path relations	hips		Unstandardized estimate	<i>S.E</i> .	р	Standardized estimate (beta coefficient)		
Self-efficacy	<	Learning atmosphere	.654	.050	***	.563		
Self-efficacy	<	Institutional support	.287	.034	***	.307		
Motivation	<	Institutional support	.373	.037	***	.349		
Motivation	<	Self-efficacy	.263	.053	***	.230		
Motivation	<	Learning atmosphere	.406	.058	***	.306		

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

Figure 2 shows that 58% variance ($R^2 = 0.58$) in motivation is explained by learning atmosphere, institutional support, and through the effect of self-efficacy. Likewise, learning atmosphere and institutional support have explained 60.0% of variance ($R^2 = 0.60$) in self-efficacy.

Finally, using bootstrapping we found that both the relationships between learning atmosphere and motivation, as well as institutional support and motivation are partially mediated by self-efficacy. In both the relationships, mediation effect is found significant at p < .001 (Table 8).

Table 8. Mediation effect in the structural model

Relationships	Direct effect	Indirect effect	Result
Institutional support> self-efficacy> motivation	.373***	.071***	Partial Mediation
Learning atmosphere> self-efficacy> motivation	.406***	.130***	Partial Mediation
** *** 0.04			

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





4.4 Student readiness for online learning

As discussed in the literature, the availability and speed of internet connection become important indicators of students' readiness. A significant portion (35.73%) of the students depend on mobile data (see Figure 3) which provides slower speed compared to the other internet connections in Bangladesh.

When asked about their preferred method for online class engagement, 22.98% of the students were in favour of pre-recorded lectures. Interestingly, 8.10% of the students do not like to participate in any form of online classes. This clearly indicates that a significant portion (22.98% and 8.18%) of the students is uncomfortable engaging in live online classes.



Figure 3. (a) Available internet connection; (b) Students' preferable mode of online classes

Further, we conducted MANOVA test to examine the effect of demographic variables on students' readiness (i.e., motivation and self-efficacy). The results of the MANOVA (Table 9) analysis suggest a statistically significant effect of all demographic variables on student readiness.

Demographic variables	Wilk's lambda (λ)	F	Hypothesis df	Error df	р	Partial eta
						squared
Gender	.987	6.560	2.0	985.0	$.001^{***}$.013
University	.910	23.873	4.0	1968.0	$.000^{***}$.046
Study Level	.983	2.130	8.0	1964.0	$.030^{*}$.009
Place of living	.951	8.338	6.0	1966.0	$.000^{***}$.025
Note $*n < 05$, *** $n < 001$						

Table 9. MANOVA analysis showing the impact of demographic variables on students' readiness

Note. **p* < .05; ****p* < .001.

We ran a separate analysis of variance (ANOVA) test to examine the statistical significance of the demographic variables on motivation and self-efficacy. We further conducted a multiple-comparison analysis (post hoc) to show exactly where the differences existed between three or more group means (Table 10).

The results indicate statistically significant impact on student readiness as follows:

- Gender showed a statistically significant impact on students' readiness for online learning, F(1, 986) = 12.96, p = .000, partial eta squared = .013, with male (M = 48.66) scoring higher than female (M = 43.38) in motivation; and F(1, 986) = 6.35, p = .012, partial eta squared = .006, with male (M = 13.69) scoring higher than female (M = 12.94) in self-efficacy.
- University type revealed a statistically significant influence on students' readiness in motivation, F(2, 985) = 45.965, p = .000, partial eta squared = .085, with public university (M = 52.37) scoring higher than international university (M = 40.55), and private university (M = 51.65) also scoring higher than international university (M = 40.55).
- Likewise, F(2, 985) = 8.065, p = .000, partial eta squared = .016, with public university (M = 14.09) scoring higher than international university (M = 12.93), and private university (M = 13.82) again scoring higher than international university (M = 12.93) in self-efficacy.
- Study level showed a statistically significant impact on students' readiness for online learning, F(4, 983) = 3.750, p = .005, partial eta squared = .015, with postgraduate students (M = 56.68) scoring higher than year 1 (M = 45.54) and year 2 (M = 45.50) students in motivation; and F(4, 983) = 2.611, p = .034, partial eta squared = .011, with postgraduate students (M = 15.22) scoring higher than year 1 (M = 13.22) students in self-efficacy.
- Living place showed a statistically significant impact on students' readiness, F(3, 984) = 7.255, p = .000, partial eta squared = .022, with village students (M = 51.95) scoring higher than both city (M = 45.64) and district town (M = 46.38) students in motivation.
- No statistically significant differences were found for living places in self-efficacy.

Demographic variables	Student readiness	Category	М	SD	df	Error	F	р	Partial eta	Post hoc
Candan	Matingtion	1 Mala	19.66	10.14	1	0.97	12.06	000***	squared	
Gender	Mouvation	1. Male	48.00	19.14	1	980	12.90	.000	.015	
	Salf	2. Pellale	43.38	3.84	1	086	635	012*	006	
	efficacy	2 Female	12.07	3.80	1	700	0.55	.012	.000	
University	Motivation	1 Public	52 37	19.80	2	985	45 965	000***	085	1>3
oniversity	Mouvation	2. Private	51.65	18.38	2	705	15.705	.000	.005	2>3
		3. International	40.55	16.74						
	Self-	1. Public	14.09	3.73	2	985	8.065	$.000^{***}$.016	1>3.
	efficacy	2. Private	13.82	3.83						2>3
	2	3. International	12.93	3.83						
Study Level	Motivation	1. Undergraduate Year 1	45.54	18.78	4	983	3.750	.005**	.015	5> 1_2
		2. Undergraduate	45.50	18.92						-, -
		3. Undergraduate	48.36	19.02						
		Year 3 4. Undergraduate	49.09	18.56						
		Year 4								
		5. Postgraduate	56.68	17.99				*		
	Self- efficacy	1. Undergraduate Year 1	13.22	3.83	4	983	2.611	.034*	.011	5>1
		2. Undergraduate Year 2	13.27	4.03						
		3. Undergraduate	13.47	3.95						
		4. Undergraduate	13.83	3.57						
		5 Postgraduate	15 22	3 78						
Place of	Motivation	1 City	45 64	18 57	3	984	7 255	000***	022	4>1
living	Monvation	2 District Town	46.28	18.09	5	201	1.200	.000	.022	2
		3 Thana Town	51.28	18.50						-
		4. Village	51.95	19.42						
	Self-	1. City	13.70	3.78	3	984	1.043	.373	.003	
	efficacy	2. District Town	13.12	3.62	-					
		3. Thana Town	13.52	3.70						
		4. Village	13.31	4.14						

T 11		1. 0	1 1 .			1. 0	1.	1 .
Table	III E test	regulte for	demographic	variables	on students'	readiness f	or online	learning
I ubic I	0.1 1051	results for	ucinographic	variables	on students	reaumess r	or onnic	icar ining

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

5. Discussion

In this study, we have developed and proposed a model for measuring engineering students' readiness for online learning in the COVID-19 situation. In developing this context-specific model, we have combined three well-known constructs: motivation, self-efficacy, and situational factors. Considering the unique situation of the COVID-19 pandemic, we have proposed the context-specific construct "situational factors" which constitute information on (i) learning atmosphere, and (ii) institutional support. We have assessed the reliability, validity, and model fit evidence of the proposed survey scale using structural equation modelling (SEM). The developed model was validated and found to be reliable for use in similar scenarios.

Situational factors, the key findings of this study, play a significant role in determining student readiness during pandemic situations, as the coefficient of determination, R^2 , indicates a high percentage of variance to explain motivation and self-efficacy (see Figure 2). In this article we argue that learning atmosphere has a pronounced impact on the extent to which engineering students are ready for online classes. Study shows that engineering students seem to be more engaged in a learning environment that offers practical-oriented, interactive, and teambased activities in an online learning environment (Kebritchi et al., 2017; Radianti et al., 2020). Boosting

students' intrinsic motivation (Ryan & Deci, 2000) by offering appropriate pedagogical modes and learning activities is likely to improve students' readiness for online classes (Hasan et al., 2016).

We also argue that institutional support plays a vital role in student motivation towards online learning and therefore their readiness learning online. If institutions provide timely IT support and a synchronized and reliable communication platform, students are likely to engage in online classes. Even if institutions provide support for online theory classes, however, more practical aspects of learning need to be included for effective online learning, especially for engineering students whose study involves practical concepts (Naji et al., 2020).

In essence, this study's most significant contribution is that it offers an instrument to measure student readiness during pandemic situations. While this study confirms the previous investigations about the influence of motivation and self-efficacy on student readiness in general (Chung et al., 2020; Hung et al., 2010; Xiong et al., 2015; Yu, 2018), additionally, it argues that situational factor is also an important phenomenon that plays a significant role on student readiness especially during a pandemic situation.

The difference between the standardized estimation values of the direct effect and the indirect effect in Table 8 confirms the mediation effect of self-efficacy between the situational factors and motivation. The higher standardized estimation values of direct effect also confirm that the situational factors' impact is higher than the impact of self-efficacy on motivation. As such, situational factors play a key role in student online readiness during a pandemic.

When a direct question was asked about the students' preferred online mode of participation, we found that approximately 30% of students did not like to engage in live online classes (see Figure 3b). This finding provides strong evidence of a low level of students' readiness for online learning during the emergency. Interestingly, students' unwillingness to engage in live online classes is commonly reported in the literature; for instance, in Handel's study (Händel et al., 2020), only 6% of students used live streaming. One potential reason for such unwillingness to attend the live classes during pandemic may be the increased number of online classes that were not usual for students, and hence difficult for them to adopt the sudden paradigm shift from full face-to-face to full online mode. Further research may explore the emerging causes of students' unwillingness to attend live classes during an emergency and normal situation.

Our data also suggest a digital inequality as a significant portion of students do not have adequate digital access in terms of internet connectivity (see Figure 3a). Using the Technology Acceptance Model (TAM) (Davis, 1989), as an investigation framework, Siron et al. (2020) argued that individuals with prior experience using computers and the Internet demonstrated higher scores in "perceived ease of use" of technology compared to new learners, and this claim is supported by the works of Lee et al. (2014) and Purnomo and Lee (2013). Because these 'at risk' or digitally-not-ready students tend to be vulnerable, a careful and deliberate instructional strategy for their online learning is required.

Our findings revealed that the differences in students' demographics (gender, university type, study level, living place) have a significant impact on student online readiness. For example, male students are likely to be more motivated and efficient than female students. This finding is supported by the study of Händel et al. (2020), however it contradicts the findings of Naji et al. (2020) and Chung et al. (2020) who reported no significant relationship between gender and student readiness. Further studies may result in better understanding of engineering students' readiness for online learning based on their gender.

Also, while differences were found among students of public, private, and international universities, the difference between public and private was not significant with respect to both motivation and self-efficacy. This may be due to some universal characteristic of students irrespective of their type of institution. Results also revealed that the junior cohort student (year 1 and year 2) is less likely to be ready than students in the senior cohort (year 3, year 4 and postgraduate). In both motivation and self-efficacy no significant differences were found among senior students. Young university students have been found to be motivated toward learning and to perform better than the senior students (Abdullah, 2011). In our case, it may be due to the pandemic that senior students become more serious about their learning to complete their study and gain employment quickly.

An interesting finding was observed when students' readiness was explored with respect to their place of living. Our data showed that village students were more motivated in online classes than city students, whereas urban students enjoyed better access to the internet than village students. The village students may believe that having less access to technology could impact negatively on their academic performance. As such, they became more motivated but also anxious about gaining access to technology and joining online classes.

5.1. Limitations and implications of this study

The survey used in this study employed convenience sampling for collecting data from the participants i.e., engineering students in Bangladesh. This sampling method can lead to unexpected or uncontrolled factors in the sample data which could potentially impact on the investigation and skew the results of the study (Emerson, 2015). However, a large sample group such as the current study may minimize the limitations posed by the convenience sampling (Etikan et al., 2016). Also, as the name indicates, convenience sampling is often used despite its limitations due to the expediency of recruiting participants (Sedgwick, 2013).

Another limitation of this study is to solely rely on self-reported survey data to measure students' readiness for online learning. We acknowledge that obtaining qualitative data through structured or semi structured interviews from some of the participants could help triangulate the data to further validate the results of this study. Future studies might consider the data triangulation approaches to gain a more comprehensive understanding of the factors that affecting students' readiness for online learning during the pandemic.

This study presented some stimulating observations which have both practical and theoretical implications for ensuring a proper learning environment for students. For example, the significance of this validated survey instrument lies in enabling institutions to assess students' readiness so they can make informed decisions about how to improve online learning, specifically, in relation to the situational factors (learning atmosphere and institutional support). These factors provide the underlying fundamentals for policy makers to design the learning context, assessment technique, etc., to prepare students for online learning. Support from educational institutes for students, in monetary or other form, would help foster a caring environment for learning too.

Informed by the insights presented, academic entities may consider establishing counselling units dedicated to supporting the students' psychological wellbeing during the pandemic as this should enhance student confidence in online learning. Institutions can consider various strategies where students with lower online readiness (i.e., motivation, self-efficacy, and situational factors) receive peer-to-peer support, guidance, or supportive intervention when they face problems or feel discouraged during the online learning. This in turn will increase student satisfaction with the education offered by their respective institutions.

Furthermore, policy makers in developing countries should consider important evidence when preparing policies for teaching in similar conditions - pandemic or otherwise - where students are required to shift to online learning due to some unwanted circumstances. Moreover, the findings will be applicable to other developing countries with similar sociodemographic conditions. Although this study focused on engineering students, some of the general findings can be applied to online learning for students from other disciplines as well.

This study revealed three key factors (motivation, self-efficacy, and situational factors) as the required conditions of student readiness for online learning. Since the current study found that computer/Internet self-efficacy and motivation for learning have direct effects on online readiness, institutions can create a simple, easy-to-use learning portal, especially where students can manage their learning resources. Such simplicity would help students feel more confident and perhaps feel less pressure to participate in the online classes.

Lessons from the study could also help teaching staff improve and customize their course teaching for such situations to improve the learning experience for students. Teaching staff should help students remain motivated since motivation is one of the important factors influencing student readiness. Students' intrinsic motivation can be increased by promoting the features of online learning i.e., creating more channels to interact with instructors and peers so that students feel a strong loving relationship among them. Rewards and extra grading can be provided to facilitate students' extrinsic motivation when students were actively engaged in online class activities, or their active participation has been recognized in any form.

6. Conclusion

The focus of this study was to investigate engineering students' readiness for online learning during the COVID-19 situation. For this, we conducted an online survey in different universities in Bangladesh and, after scrutiny, selected 988 responses out of 1038 initial responses. We collected engineering students' opinions on factors that influence students' readiness for online learning. Our study proposed a new model to measure student readiness for online learning considering the context of the COVID-19 situation. The reliability, convergent and discriminant validity of the proposed model was tested using EFA and CFA methods. Twenty-two items were removed from the original 52 items to achieve composite reliability greater than 0.7. Our study suggests that besides motivation and self-efficacy, situation and context-specific factors influence students' readiness for online learning. It is evident from the findings that students are not ready yet for online learning. Besides the usual student unwillingness (Händel et al., 2020), our study shows that student readiness towards online learning is hindered by digital inequality in a developing country due to lack of experience and access to relevant technologies. In developing countries like Bangladesh, the penetration of internet connectivity is widely varied; hence students lived in rural areas are seemly less accessed to the internet.

Moreover, it becomes more severe during the pandemic as students' need to move their home areas to stay with families. The proposed model can be helpful to improve the student learning experience in emergencies and address potential issues related to student online readiness. A longitudinal study may be performed in future to detect any changes in the relationship of the factors considered in this study. We also plan to extend this study by broadening the demographic distribution to include participants from different disciplines.

Acknowledgement

The study is approved by the office of Research, Extension, Advisory Services and Publications (REASP), Islamic University of Technology. We also acknowledge the support provided by the School of Business and Digital Technologies, Manukau Institute of Technology for proofreading the document.

References

Abdullah, A.-M. (2011). Factors affecting business students' performance in Arab Open University: The Case of Kuwait. *International Journal of Business and Management*, 6(5), 146-155. https://doi.org/10.5539/ijbm.v6n5p146

Ahmed, M. (2020). Managing schools, learning and student wellbeing during Covid-19. *The Daily Star*. https://www.thedailystar.net/opinion/news/managing-schools-learning-and-student-wellbeing-during-covid-19-1893556

Allam, S. N. S., Hassan, M. S., Sultan, R., Mohideen, A. F. R., & Kamal, R. M. (2020). Online distance learning readiness during Covid-19 outbreak among undergraduate students. *Journal of Academic Research in Business and Social Sciences*, 10(5), 642-657.

Arthur-Nyarko, E., Agyei, D. D., & Armah, J. K. (2020). Digitizing distance learning materials: Measuring students' readiness and intended challenges. *Education and Information Technologies*, 25, 2987–3002.

Beaunoyer, E., Dupéré, S., & Guitton, M. J. (2020). COVID-19 and digital inequalities: Reciprocal impacts and mitigation strategies. *Computers in Human Behavior*, 111, 106424. https://doi.org/10.1016/j.chb.2020.106424

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, 88(3), 588-606. https://doi.org/10.1037/0033-2909.88.3.588

Bisht, R. K., Jasola, S., & Bisht, I. P. (2020). Acceptability and challenges of online higher education in the era of COVID-19: a study of students' perspective. *Asian Education and Development Studies*. https://doi.org/10.1108/AEDS-05-2020-0119

Blunch, N. (2008). Introduction to structural equation modeling using IBM SPSS statistics and AMOS. Sage.

Bonanati, S., & Buhl, H. M. (2021). The Digital home learning environment and its relation to children's ICT self-efficacy. *Learning Environments Research*, 1-21. https://doi.org/10.1007/s10984-021-09377-8

Chang, C.-Y., Kao, C.-H., & Hwang, G.-J. (2020). Facilitating students' critical thinking and decision making performances: A Flipped classroom for neonatal health care training. *Educational Technology & Society*, 23(2), 32-46.

Cho, M.-H., & Tobias, S. (2016). Should instructors require discussion in online courses? Effects of online discussion on community of inquiry, learner time, satisfaction, and achievement. *International Review of Research in Open and Distributed Learning*, *17*(2), 123-140.

Chung, E., Subramaniam, G., & Dass, L. C. (2020). Online learning readiness among university students in Malaysia amidst Covid-19. *Asian Journal of University Education*, *16*(2), 45-58.

Cox, A. M. (2017). The Experience of learning atmosphere in the learning commons. In *iConference 2017 Proceedings* (pp. 333-339). iSchools. https://doi.org/10.9776/17201

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

DeVillis, R. F. (2016). Scale development: Theory and applications (4th ed.). Sage.

Emerson, R. W. (2015). Convenience sampling, random sampling, and snowball sampling: How does sampling affect the validity of research? *Journal of Visual Impairment & Blindness, 109*(2), 164-168.

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics*, 5(1), 1-4.

Fernandes, N. (2020). Economic effects of coronavirus outbreak (COVID-19) on the world economy. *IESE Business School*, WP-1240-E. http://dx.doi.org/10.2139/ssrn.3557504

Field, A. (2013). Discovering statistics using IBM SPSS statistics. Sage.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, 18(1), 39-50.

Galindo-Dominguez, H. (2021). Flipped classroom in the educational system: Trend or effective pedagogical model compared to other methodologies? *Educational Technology & Society*, 24(3), 44-60.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2010). *Multivariate data analysis: A Global Perspective*. Pearson Prentice Hall.

Hair, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2016). A Primer on partial least squares structural equation modeling (PLS-SEM). Sage Publications.

Han, F., & Ellis, R. A. (2021). Predicting students' academic performance by their online learning patterns in a blended course: To what extent is a theory-driven approach and a data-driven approach consistent? *Educational Technology & Society*, 24(1), 191-204.

Händel, M., Stephan, M., Gläser-Zikuda, M., Kopp, B., Bedenlier, S., & Ziegler, A. (2020). Digital readiness and its effects on higher education student socio-emotional experiences in the context of COVID-19 pandemic. *Journal of Research on Technology in Education*. https://doi.org/10.1080/15391523.2020.1846147

Hasan, H., Linger, H., Chen, A., Lu, Y., & Wang, B. (2016). Enhancing perceived enjoyment in social games through social and gaming factors. *Information Technology & People*, 29(1), 99-119. https://doi.org/10.1108/ITP-07-2014-0156

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A New criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the academy of marketing science*, 43(1), 115-135.

Hossain, M. S., Ferdous, S., & Siddiqee, M. H. (2020). Mass panic during Covid-19 outbreak- A Perspective from Bangladesh as a high-risk country. *Journal of Biomedical Analytics*, 3(2), 1-3.

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

Huang, R., Liu, D., Tlili, A., Yang, J., & Wang, H. (2020). Handbook on facilitating flexible learning during educational disruption: The Chinese experience in maintaining undisrupted learning in COVID-19 Outbreak. Smart Learning Institute of Beijing Normal University.

Hung, M.-L., Chou, C., Chen, C.-H., & Own, Z.-Y. (2010). Learner readiness for online learning: Scale development and student perceptions. *Computers & Education*, 55(3), 1080-1090.

Hutcheson, G. D., & Sofroniou, N. (1999). The Multivariate social scientist: Introductory statistics using generalized linear models. Sage.

Ioannou, M., & Ioannou, A. (2020). Technology-enhanced embodied learning: Designing and evaluating a new classroom experience. *Educational Technology & Society*, 23(3), 81-94.

Irani, T. A., Wilson, S. B., Slough, D. L., & Rieger, M. (2014). Graduate student experiences on- and off-campus: Social connectedness and perceived isolation. *International Journal of E-Learning & Distance Education*, 28(1). http://www.ijede.ca/index.php/jde/article/view/856

Joksimović, S., Gašević, D., Kovanović, V., Riecke, B. E., & Hatala, M. (2015). Social presence in online discussions as a process predictor of academic performance. *Journal of Computer Assisted Learning*, *31*(6), 638-654. https://doi.org/10.1111/jcal.12107

Kebritchi, M., Lipschuetz, A., & Santiague, L. (2017). Issues and challenges for teaching successful online courses in higher education: A Literature review. *Journal of Educational Technology Systems*, *46*(1), 4-29.

Khine, M. S., Fraser, B. J., & Afari, E. (2020). Structural relationships between learning environments and students' noncognitive outcomes: secondary analysis of PISA data. *Learning Environments Research 2020, 23*(3), 395–412. https://doi.org/10.1007/S10984-020-09313-2

Khtere, A. R., & Yousef, A. M. F. (2021). The Professionalism of online teaching in Arab universities: Validation of faculty readiness. *Educational Technology & Society*, 24(3), 1-12.

Kline, R. B. (2016). Principles and practice of structural equation modeling (4th ed.). The Guilford Press.

Kokoç, M., Akçapınar, G., & Hasnine, M. N. (2021). Unfolding students' online assignment submission behavioral patterns using temporal learning analytics. *Educational Technology & Society*, 24(1), 223-235.

Lai, H.-J. (2011). The Influence of adult learners' self-directed learning readiness and network literacy on online learning effectiveness: A Study of civil servants in Taiwan. *Educational Technology & Society*, 14(2), 98-106.

Lee, Y.-H., Hsiao, C., & Purnomo, S. H. (2014). An Empirical examination of individual and system characteristics on enhancing e-learning acceptance. *Australasian Journal of Educational Technology*, *30*(5).

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological methods*, 1(2), 130-149.

Manganello, F., Falsetti, C., & Leo, T. (2019). Self-regulated learning for web-enhanced control engineering education. *Educational Technology & Society*, 22(1), 44-58.

Miglani, A., & Awadhiya, A. K. (2017). Mobile learning: Readiness and perceptions of teachers of Open Universities of Commonwealth Asia. *Journal of Learning for Development*, 4(1), 58-71.

Monjur, M. R., & Hassan, M. Z. (2020). Early phases of COVID-19 management in a low-income country: Bangladesh. *Infection Control & Hospital Epidemiology, 1*. https://doi.org/10.1017/ice.2020.147

Neuwirth, L. S., Jović, S., & Mukherji, B. R. (2020). Reimagining higher education during and post-COVID-19: Challenges and opportunities. *Journal of Adult and Continuing Education*, 1-16. https://doi.org/10.1177/1477971420947738

Pamungkas, D. K. (2019). Comparing learning motivation and student achievement using various learning models. *Classroom Action Research Journal*, 2(2), 64-70.

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management*, 12(4), 531-544. https://doi.org/10.1177/014920638601200408

Purnomo, S. H., & Lee, Y.-H. (2013). E-learning adoption in the banking workplace in Indonesia: An Empirical study. *Information Development*, 29(2), 138-153.

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A Systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education, 147*, 103778. https://doi.org/10.1016/j.compedu.2019.103778

Richardson, M., Sheeks, G., Waller, R. E., & Lemoine, P. A. (2021). Pursuit of online services for online students. In D. M. Velliaris (Ed.), *Research Anthology on Developing Effective Online Learning Courses*. https://doi.org/10.4018/978-1-7998-8047-9.ch092

Rohatgi, A., Scherer, R., & Hatlevik, O. E. (2016). The Role of ICT self-efficacy for students' ICT use and their achievement in a computer and information literacy test. *Computers & Education*, *102*, 103–116. https://doi.org/10.1016/j.compedu.2016.08.001

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68-78. https://doi.org/10.1037/0003-066X.55.1.68

Schumacker, R. E., & Lomax, R. G. (2010). A Beginner's guide to structural equation modeling (3rd ed.). Lawrence Erlbaum Associates.

Sedgwick, P. (2013). Convenience sampling. British Medical Journal, 347:f6304. https://doi.org/10.1136/bmj.f6304

Siron, Y., Wibowo, A., & Narmaditya, B. S. (2020). Factors affecting the adoption of e-learning in Indonesia: Lesson from Covid-19. *Journal of Technology and Science Education*, *10*(2), 282-295.

Trespalacios, J., Uribe-Flórez, L., Lowenthal, P. R., Lowe, S., & Jensen, S. (2021). Students' perceptions of institutional services and online learning self-efficacy students' perceptions of institutional services and online learning self-efficacy. *American Journal of Distance Education*. https://doi.org/10.1080/08923647.2021.1956836

Twenge, J. M., Spitzberg, B. H., & Campbell, W. K. (2019). Less in-person social interaction with peers among U.S. adolescents in the 21st century and links to loneliness. *Journal of Social and Personal Relationships*, *36*(6), 1892-1913. https://doi.org/10.1177/0265407519836170

Ullah, M. N., Biswas, B., & Miah, M. (2021). Assessing institutional support to online education at tertiary level in Bangladesh coping with Covid-19 pandemic: An Empirical study. *Journal of Digital Educational Technology*, 2(1), ep2204. https://doi.org/10.21601/jdet/11735

Wang, W. T., & Lin, Y.-L. (2021). The Relationships among students' personal innovativeness, compatibility, and learning performance: A Social cognitive theory perspective. *Educational Technology & Society*, 24(2), 14-27.

Xiong, Y., So, H.-J., & Toh, Y. (2015). Assessing learners' perceived readiness for computer-supported collaborative learning (CSCL): A Study on initial development and validation. *Journal of Computing in Higher Education*, 27(3), 215-239.

Yeasmin, S., Banik, R., Hossain, S., Hossain, M. N., Mahumud, R., Salma, N., & Hossain, M. M. (2020). Impact of COVID-19 pandemic on the mental health of children in Bangladesh: A Cross-sectional study. *Children and Youth Services Review*, *117*, 105277. https://doi.org/10.1016/j.childyouth.2020.105277

Yu, T. (2018). Examining construct validity of the Student Online Learning Readiness (SOLR) instrument using confirmatory factor analysis. *Online Learning*, 22(4), 277-288.

Yu, T., & Richardson, J. C. (2015). An Exploratory factor analysis and reliability analysis of the Student Online Learning Readiness (SOLR) instrument. *Online Learning*, 19(5), 120-141.

Zheng, L., Bhagat, K. K., Zhen, Y., & Zhang, X. (2020). The Effectiveness of the flipped classroom on students' learning achievement and learning motivation: A Meta-analysis. *Educational Technology & Society*, 23, 1176-3647.

Appendix

The initial "Student Online Learning Readiness" Survey

A. Motivation

Sub-constructs	Items
Interest	MI1. I think I enjoy learning very much in online environment.
	MI2. I think learning in online environment is a boring activity [*] .
	MI3. I would describe learning activity in online environment as very interesting.
	MI4. I think online learning activity is quite enjoyable.
	MI5. I am open to accept the online environment for my learning.
	MI6. I like to work with my classmates in an online environment.
	MI7. I like to work with my teachers in an online environment.
Perceived	MU1. I believe it is effective to learn in online classes.
Usefulness	MU2. I believe online classes can help my learning.
	MU3. I believe online classes help me to learn more complex topics than face-to-face
	classroom.
	MU4. I believe online classes allow many opportunities for discussion and sharing ideas among my classmates.
	MU5. I would be willing to learn in online classes again because it has some value to me.
	MU6. I think online learning is important because it can improve my learning.
	MU7. I believe online learning activity could be beneficial to me.
Reinforcement	MR1. Through online classes, I hope to achieve a good grade for the courses I attend.
	MR2. I hope my teachers and classmates will praise me if I can perform good in online
	classes.
	MR3. I hope my attendance in online classes will improve my course grade.
	MR4. I hope online classes will have a positive impact in my career.
Connectedness/	MC1. I like to connect with my teachers and classmates in the online learning environment.
Relatedness	MC2. I feel like I can trust my teachers in the online learning environment.
	MC3. I prefer not to interact with my teachers and classmates in the online learning
	environment in future [*] .
	MC4. I feel disconnected from my teachers and classmates in the online learning
	environment [®] .
	MC5. I feel close to my teachers and classmates in the online learning environment.
	MC6. I feel I could develop friendship with my teachers and other students in the online
	learning environment.
	MC/. I would like to interact with my teachers and classmates more often in the online
	learning environment.
**. *	MC8. I feel I could develop a good bonding with others through online learning environment.
Item needs rever	se coding

B. Self-efficacy

Technology	TC1. I feel confident in performing the basic functions of technology used in online learning.
Competency	TC2. I feel confident in my knowledge and skills of how to manage software for online
	learning.
	TC3. I feel confident in using the internet to find or gather relevant information for learning.
	TC4. I feel competent at integrating computer technologies into my learning activities.
Social	SC1. I feel confident to ask questions to my teachers in online classes.
Competency	SC2. I feel confident to seek help from my teachers when needed.
	SC3. I feel confident to timely inform my teachers when unexpected situations arise.
	SC4. I feel confident to express my opinions to teachers respectfully.
	SC5. I feel confident to initiate discussions with my teachers in online environment.
	SC6. I feel confident to respect other students' social actions in online environment.
	SC7. I feel confident to apply different social interaction skills depending on situations.
	SC8. I feel confident to initiate social interaction with classmates.
	SC9. I feel confident to work in groups in online environment.
	SC10. I feel confident to develop friendship with my classmates in online environment.

C. Situational Factors

Learning	LA1. I think my living environment is supportive to study in online environment.
atmosphere	LA2. I think I can effectively study from my living place.
	LA3. I think my family members around me are helpful for my online study.
	LA4. I think it is difficult to study online from the place where I am living [*] .
Institutional	IS1. I believe my institution is supportive for my online study.
Support	IS2. I believe I can get the necessary help from my institution to study online.
	IS3. I believe my institution makes necessary arrangements for effective online learning.
	IS4. I believe my institution can provide a favourable environment for my online study.
Digital access	DA1. I believe I have the necessary devices to participate in online classes.
	DA2. I believe I can afford the cost of internet to participate in online classes.
	DA3. I believe the internet connection and speed is reliable enough for the online classes.
	DA4. I think I do not have enough resources to study online [*] .
**. 7	1.

*Item needs reverse coding

Supplementary dataset

Available from Mendeley <u>data repository</u>.

The Impact of Teaching Simulation Using Student Chatbots with Different Attitudes on Preservice Teachers' Efficacy

Donggil Song¹, Eun Young Oh² and Hyeonmi Hong^{3*}

¹Sam Houston State University, USA // ²Rice University, USA // ³Jeju National University, South Korea // donggil.song@gmail.com // oh@rice.edu // hongspiderweb@gmail.com *Corresponding author

(Submitted May 14, 2021; Revised October 31, 2021; Accepted November 29, 2021)

ABSTRACT: The aim of this study is to evaluate the effects of a teaching simulation activity that uses a chatbot on preservice teachers' efficacy. Forty-six preservice teachers were asked to teach the chatbot the topic of school violence and how to handle it. They were assigned to one of three groups: Teaching a chatbot whose attitude was impolite, polite, or ordinary. The participants completed a teacher efficacy test at the pretest and posttest. The results show that the participants who taught the ordinary chatbot significantly increased their teacher efficacy levels. However, an Analysis of Covariance shows that the posttest scores of student engagement were not significantly different due to the group setting. The data of participants' conversations with the chatbots and the participant interview data revealed that the ordinary group had more opportunities to increase their mastery experiences than the other groups did. It can be suggested that designing virtual students with ordinary and regular attitudes and behaviors seems appropriate to provide preservice teachers with teaching opportunities to increase their teaching efficacy levels.

Keywords: Teaching simulation, Virtual agent, Chatbot, Teacher efficacy

1. Introduction

It is imperative to increase levels of teacher efficacy because it is directly linked to teacher development, such as instructional efforts, content knowledge (Menon & Sadler, 2016), and their students' academic performance and aspiration (Muijs & Reynolds, 2002; Zee & Koomen, 2016). Teaching experience is considered a powerful influence on teachers' efficacy in and perception of their competence, determining the extent to which teachers will persevere in their classroom (Klassen & Chiu, 2010). Unfortunately, preservice teachers are rarely exposed to teaching opportunities; in most cases, they have a limited amount of student-teaching practicum. Therefore, more opportunities for teaching should be offered during teacher education programs.

Previous research has highlighted the importance of teaching experience for preservice teachers by focusing on microteaching (Arsal, 2015), teaching simulation (Polack et al., 2017), and student-teaching (Fives et al., 2007). However, managing these teaching activities is demanding, and face-to-face implementation is time-intensive. Mostly due to these practical issues, little attention was paid to research on how to improve preservice teachers' efficacy through teaching activities. To increase teaching efficacy and simultaneously resolve practical difficulties, it would be worth exploring the use of teaching simulation. Still, how to simulate teaching experiences to overcome the current challenges of teaching activities is under-investigated.

The aim of this study is to explore the potentials of chatbots to provide preservice teachers with teaching practice opportunities in a learning environment. We aim to clarify the possibilities and challenges of a teaching simulation program that pays particular attention to a conversational approach that uses a virtual agent system (also called chatbot), which is expected to give preservice teachers quasi-experiences of teaching. Specifically, we focus on the effects of the chatbot's different attitudes on preservice teachers' efficacy.

2. Literature review

This study is grounded in Bandura's (1977) social cognitive theory of behavior with a focus on self-efficacy. Bandura (1977) introduced the concept of self-efficacy beliefs, which is an assessment of one's capabilities to attain the desired performance goal. Self-efficacy is a significant component in the human agency as it influences people's aspirations and behaviors, including choice of tasks, effort, and persistence (Bandura, 1986). Since the self-efficacy theory was applied in the education realm, a considerable number of studies have been conducted to investigate how self-efficacy is related to teaching performance and student achievement (Bautista & Boone, 2015; Zee & Koomen, 2016), which has evolved into the concept of teacher efficacy. In many studies,

Tschannen-Moran and Woolfolk Hoy's (2001) definition of teacher efficacy was introduced as "a teacher's judgment of his or her capabilities to bring about desired outcomes of student engagement and learning, even among those students who may be difficult or unmotivated" (p. 783). In this definition, there is an assumption of teachers' beliefs in their ability to positively affect student learning and behavior (Putman, 2012).

Higher levels of teacher efficacy led to more time investment in teaching and greater levels of aspiration (Tschannen-Moran & Woolfolk Hoy, 2001), teacher development (Menon & Sadler, 2016), student performance (Kim & Seo, 2018; Muijs & Reynolds, 2002), and student motivation (Appleton & Kindt, 2002; Lazarides et al., 2018). Teacher efficacy appears to be a belief that affects both teaching and student learning in a significantly positive way. The low level of teacher efficacy causes negative issues, such as a lack of teaching preparation (Mulholland & Wallace, 2001), job dissatisfaction (Klassen & Chiu, 2010; Perera et al., 2018), and emotional exhaustion (Skaalvik & Skaalvik, 2016). These negative aspects predict greater teacher turnover intent (Ryan et al., 2017). A high rate of teacher attrition has been one of the severe issues that cause teacher shortage problems.

2.1. Teacher efficacy promotion

Bandura (1986) categorized the sources of self-efficacy development into mastery experiences, vicarious experiences, social persuasion, and physiological/emotional factors. In the teacher education field, it can be understood as follows. First, mastery experiences are achieved by conducting teaching. Second, preservice teachers have vicarious experiences through modeling others' teaching. Preservice teachers observe in-service teachers' classrooms, other media, or their own teaching through recorded videos. Third, social persuasion means that novice teachers utilize others' feedback. Last, physiological and emotional factors are related to teachers' stress reactions and negative tendencies. Among them, mastery experiences have been identified as the most effective method (Tschannen-Moran & Woolfolk Hoy, 2007). It was reported that beginning elementary teachers showed their improved teacher efficacy through the achievement of mastery experiences of teaching (Mulholland & Wallace, 2001). It can be argued that teaching in a classroom could be essential for teachers because those experiences provide authentic evidence of their accomplishment from successful teaching (Pfitzner-Eden, 2016). Because mastery experiences originate from teaching accomplishments, preservice teachers would not have enough opportunities for improving teaching efficacy if there was a lack of teaching opportunities in their programs, such as not enough practicum and student-teaching courses. Novice and preservice teachers have relatively lower levels of self-efficacy than experienced career teachers do, which could be due to the lack of teaching experience, not because of their innate ability (Gordon & Debus, 2002).

2.2. Mastery experiences

Given the importance of mastery experiences, one of the opportunities to effectively increase efficacy is studentteaching. Cantrell et al. (2003) found that the amount of time spent in student-teaching as mastery experiences were positively correlated with their teacher efficacy. Pfitzner-Eden (2016) examined the teacher efficacy of two cohorts of preservice teachers. After the practicum, two factors of teacher efficacy (i.e., classroom management and instructional strategy) were increased in the group of preservice teachers at the beginning stage. The results of the advanced preservice teachers also showed the increased teacher efficacy of classroom management. Fives et al. (2007) explored preservice teachers' engagement in practicum and their perception of teacher burnout. The results of preservice teachers' student-teaching experience indicate that their teacher efficacy was significantly increased, and burnout symptoms were decreased over time. The researchers argue that student teaching provides mastery experiences, and at the same time, ameliorates preservice teachers' feelings of burnout. Thus, preservice teachers' mastery experiences in student-teaching seem essential to improve their teacher efficacy.

2.3. Students' attitudes and behaviors

Despite the benefits of student-teaching in general, effective student-teaching environments and contexts have been under-explored. Specifically, it seems that students' attitudes and behaviors have an impact on preservice teachers' efficacy (Dicke et al., 2014; Kokkinos et al., 2005). One of the critical causes of a high level of teacher stress is student misbehavior and disengagement, which is the biggest concern of novice teachers who lack teaching experience and coping strategies (Dicke et al., 2014). Students' disruptive behaviors and hostile attitudes evoke teachers' unfavorable and negative emotions significantly (Kokkinos et al., 2005), which negatively affect preservice teachers' intrinsic motivation to teach, goals towards their profession, and teaching efficacy (Sutton & Wheatley, 2003).

It is imperative to provide preservice teachers with opportunities to learn how to handle the different attitudes of students. In Kokkinos et al. (2005), over five hundred primary school teachers and preservice teachers completed the survey that asked their appraisals of students' undesirable behaviors. The results implicate that more teacher training should be designed to increase the awareness of a broader range of students' undesirable behaviors, provide how to cope with student difficulties, and enhance teacher efficacy in handling challenging and disruptive behavior. Kim and Cho (2014) examined the status of preservice teachers in their teacher education program and their expectation of "reality shock," which refers to a huge discrepancy between the ideal expectations of teaching and the reality of the school environment and student attitudes. The results show that the expectation of reality shock varies depending on the status in their program, which has three stages: (1) before declaring their teaching major: the highest level of expectation of future reality shock, (2) after having decided on their major: the significantly decreased expectation of reality shock, and (3) after their practicum: the high level of expectation of future reality shock. This could mean that as preservice teachers are exposed to a classroom in their student-teaching, they clearly understand that the teaching context is not only about delivering content knowledge but about coping with difficult students.

For successful mastery experiences, preservice teachers should be able to have enough opportunities to teach students with different attitudes and behaviors, including problematic and disruptive ones. One or two courses of student-teaching practicum might not fully provide mastery experiences opportunities of teaching different types of students. Few studies examined how student attitudes and behaviors are related to preservice teachers' efficacy. In addition, how to support preservice teachers' mastery experiences was under-investigated.

Along with the lack of research on the identification of contextual factors (i.e., students' attitudes and behaviors), there are practical issues when we attempt to provide sufficient student-teaching opportunities. Due to the lack of opportunities of teaching different attitudes' students, preservice teachers might not be fully prepared for their teaching, such as the absence of a direct link between instructional goals and assessment, failure to write observable instructional goals, and the lack of foundational concepts as validity and reliability when assessing students (Campbell & Evans, 2000). Besides, there is a disconnect between the campus-based portion of teacher education programs and student-teaching. This might be because practicum courses are not considered as a valued activity, and supervising practicum courses has been treated as an overload (Zeichner, 2002). To resolve these practical limitations, it requires an effort of educators to propose alternative ways for successful mastery experiences.

2.4. Chatbots

Our approach is to provide preservice teachers with sufficient teaching practice opportunities using learning technology. In the field of education, conversational virtual agents, called AI (artificial intelligence) agents or chatbots, have been investigated. Chatbots are computer programs that communicate in human language with their users. The system can conduct interaction activities through communication with the user by simulating human-like dialog patterns and behavior. In 1966, Joseph Weizenbaum developed one of the first chatbots, ELIZA, which simulates a therapist's role in clinical treatment situations. Since this breakthrough, chatbots have evolved and been utilized in a variety of areas, such as marketing, customer support, e-commerce, banking, and healthcare.

Chatbots have been adopted in education. One type is Intelligent Tutoring Systems, which have shown effectiveness in improving learning achievements specifically in science, technology, engineering, and mathematics fields (VanLehn, 2011). Winkler and Soellner (2018) reviewed that chatbots have been effectively used for medical education and therapy, language learning, feedback systems, and motivation and self-efficacy supporters. Furthermore, their findings show that chatbots have the potential to improve learners' affective, cognitive, and metacognitive learning gains (Winkler & Soellner, 2018). There are cases that chatbots are used for educational purposes (e.g., Oh et al., 2019; Song & Kim, 2020). Abbasi and Kazi (2014) investigated the use of a chatbot as a question retrieval tool to support students in solving programming questions. Seventy-two undergraduate students were randomly assigned to one of two groups: Google Group (who searched for information using Google search engine to solve questions) and Agent Group (who asked questions to the conversational agent to retrieve information for problem-solving). The results show that the learning outcomes of Agent Group were significantly higher than those of Google Group. Chatbots can be used to scaffold students' positive habit development. Kreynin et al. (2019) measured whether chatbots may assist undergraduate students to develop a positive habit such as reflective journaling. The researchers used four versions of a chatbot that enables effective journaling via text messages. The results showed that chatbots can be effective tools in scaffolding positive habit development for undergraduate students (Kreynin et al., 2019). However, chatbots have not been thoroughly investigated in the teacher education field.

We argue that chatbots would lead to an in-depth research investigation of the effectiveness of teaching simulation and the attitudes/behaviors of agents (i.e., virtual students). Still, the use of chatbots to offer student-teaching simulation opportunities is in its infancy. There is little knowledge of what roles AI technology might play in preservice teacher education.

2.5. This study

We aim to investigate the effects of teaching simulation through a chatbot on preservice teachers' efficacy levels. Two research questions guided the study:

- RQ1. To what extent, does the preservice teachers' efficacy change by teaching a student chatbot?
- RQ2. Are there differences in preservice teachers' efficacy associated with teaching the different attitudes of chatbots (i.e., impolite, polite, and ordinary)?
- RQ3. What are the preservice teachers' responses and experiences in teaching chatbots with different attitudes?

3. Methods

3.1. Participants and context

A sample of 46 students (14 males and 32 females) from Teachers College at a mid-sized public university in South Korea provided data for this study. The research information was explained to the participants. They are undergraduate students (senior) majoring in elementary education, who had a one-time experience of student-teaching a year prior to the intervention of this study. They were randomly assigned to one of three groups: Teaching the agent whose attitude was (1) impolite (Impolite Group, N = 16), (2) polite (Polite Group, N = 16), and (3) ordinary (Ordinary Group, N = 15).

3.1.1. Teaching subject

Over the last decade, school violence has been recognized as a serious problem at the research site. Increasing school violence problems lead to the increased pressure on teachers to prevent them in the classroom, and teachers should play a significant role in tackling school violence (Troop-Gordon, 2015; Yoon et al., 2016). Specifically, preservice teachers should be well prepared to teach anti-bullying issues to their future students. For these reasons, anti-bullying was chosen as a topic for the teaching simulation in this study. Before the teaching task, participants were asked to prepare their anti-bullying teaching sessions for the teaching simulation activity.

3.1.2. System design and development

A teaching simulation program was designed to support conversations between participants and a chatbot. The system was built on an existing chatbot framework that was designed to support users' interaction with the agent (Song et al., 2017; Oh et al., 2019; Song & Kim, 2020). When a participant accesses the agent system through their Web browser, the agent greets the participant, and the teaching session begins. The participant's role is the teacher, and the agent takes the student's role.

For the implementation, the participants were asked to teach the concept of school violence. The teaching method is individual text-based chatting, which is similar to social media message apps. Participants are textbased chatting with a chatbot to teach. The agent initiates conversation with greetings and questions, and a participant teaches the agent. The system analyzes a pattern of what the participant typed (e.g., a positive answer to the agent's question, a follow-up question) and responds to the participant accordingly. In the conversation, attention was specifically given to the role of the agent. Using machine learning techniques, each chatbot agent was trained to have a specific attitude. Depending on their group, the participants taught the agent whose attitude was impolite, polite, or ordinary. The impolite agent plays in distracting participants' conversation and showing disrespect and discourtesy to them, whereas the polite agent in revealing his understanding and showing gratitude to the group of preservice teachers. The ordinary agent responds to the participants ordinarily, not showing any specific negative or positive emotion and attitude. The agent was designed as gender-neutral and the participants could not recognize the agent's gender. The agent's age was not specifically addressed in the program, but it was designed as 5th – 6th graders.

The agents' responses were designed (i.e., machine learning training) from the results of multiple interviews with local teachers and students at the design and development stage before several pilot tests. The development team collected all the possible students' reactions and responses depending on their attitudes and added them into the system database. The development team collected possible sentences that in-service teachers used in their antibullying interventions at school and listed them as the participants' anticipated teaching. Each agent's attitude was designed based on formal/informal interviews and conversations with the teachers and students. At the development stage, the development team asked students to come up with possible responses and reactions to each teaching sentence from the list. Then, each sentence had multiple students' responses, which were categorized into impolite, polite, or ordinary responses, and inserted into the database by the development team. This categorization of attitude was conducted by the research team first and was cross-checked by the participanting teachers. When a participant starts teaching the chatbot, using classification techniques of machine learning, the system categorizes the participant's sentence into one of the anticipated teaching sentences that were collected through the interviews with the local teachers. Depending on the type of teaching sentence, the agent was designed to respond to the participant appropriately using the database of possible students' reactions and responses.

3.2. Data collection and analysis

3.2.1. Pretest and posttest

The participants completed a teacher efficacy test at the pretest and posttest. Teachers' Sense of Efficacy Scale developed by Tschannen-Moran and Woolfolk Hoy (2001) was used to assess the efficacy of participants due to its recognized acceptance in the research field and its validation with preservice teachers. The instrument consists of 24 items, assesses along a 9-point continuum with anchors at 1 (Nothing) to 9 (A Great Deal). Considering the survey type that the participants of this study were familiar with, each question item was modified from a question sentence to an assertive sentence, and the measure was tweaked to a five-step Likert scale: Strongly Disagree (1) to Strongly Agree (5). The teacher efficacy instrument includes three 8-item sub-dimensions: Student Engagement, Instructional Strategies, and Classroom Management. A reliability test was conducted for the test items from the pretest. The reliability of the full 24-item scale was .85. Reliabilities for the teacher efficacy subscales (i.e., student engagement, instructional strategies, and classroom management) were .70, .74, and .74, respectively. The student engagement dimension indicates a relatively low but acceptable level (conventionally greater than .6; Wiersma & Jurs, 2005) of internal consistency, and the other dimensions show acceptable levels of internal consistency.

3.2.2. Interview and written discourse data

Since the two research questions include "how" questions, interview and written discourse data were analyzed to complement quantitative results. The quantitative results would give us the amount of student attitude change, and qualitative analysis might answer how this change happened. After the intervention, four volunteered participants from each group were interviewed. The interview questions asked participants' experience of the teaching simulation focusing on the agent's behavior and attitude, their emotions during the text-based conversation, and feeling about teacher efficacy. In addition, participants' conversation data was collected through the agent system. The system collected each participant's session identification number, timestamp, and conversation text. The student interview and written discourse data were qualitatively analyzed.

We followed a thematic analysis method (Braun & Clarke, 2006). The qualitative data were coded to determine major themes and categories that would emerge from the data through a process of reading and rereading the data. The transcript was independently coded by the first author and the second author. The calculated inter-rater reliability ranged from 90% to 95%, averaging 92.7%. These disagreements were further discussed and resolved by the coders. The meaningful statements regarding teaching experience were initially highlighted from the interview and written discourse data. Then, the coding process was conducted by marking the segments of data with descriptive words or category names, such as "the impolite agent," "the difficulty levels of teaching," "interaction," and "teaching practice." Then, the codes were grouped into sub-categories (e.g., teaching conversation, persuasion effort, off-topic conversations). The related subcategories were integrated by comparing and contrasting the properties of each sub-category, and continuously refining and collapsing the sub-categories as stronger themes emerged. Similar ideas were summarized into statements to present the common themes and

insights from the participants' teaching simulation experience. The common issues were addressed using the analyzed themes.

4. Results

4.1. RQ1

The normality of the data was checked through the Shapiro–Wilk test. All test scores did not deviate significantly from the normal distribution (see Table 1). Means and standard deviations for teacher efficacy are presented in Table 2. A paired sample *t*-test was conducted in each group to examine the difference between the pretest and posttest scores of the teacher efficacy test. As shown in Table 2, using a nominal alpha value of .05, a significant mean difference between pretest and posttest was found for the total score of Ordinary Group (*t* = - 4.47, *p* = .001, Cohen's *d* = .50).

Table 1. Tests of normality on efficacy test					
	Impolite Group Polite Group Ordinary Group Total				
	Shapiro–Wilk (<i>p-value</i>) Shapiro–Wilk (<i>p-value</i>) Shapiro–Wilk (<i>p-value</i>)		Shapiro–Wilk (<i>p-value</i>)		
	(df = 16)	(df = 15)	(df = 15)	(df = 46)	
Pre	.97 (.76)	.95 (.50)	.96 (.76)	.98 (.72)	
Post	.95 (.47)	.98 (.93)	.95 (.45)	.99 (.80)	

Table 2. Mean scores and	1 standard deviations of	n efficacy test for the	participants and <i>t</i> -test results
1 dole 2. filedii beores dii	a blandara ac rianono or	i ennede y test for the	purchespunds und r cost results

Total	Impolite Group Mean (SD)	Polite Group Mean (SD)	Ordinary Group Mean (SD)
	(N = 16)	(N = 15)	(<i>N</i> = 15)
Pre	3.39 (.31)	3.64 (.27)	3.52 (.40)
Post	3.50 (.34)	3.58 (.54)	3.73 (.44)
<i>t-value (p-value)</i>	-2.09 (.054)	.44 (.665)	-4.47 (.001)*

Note. **p* < .05.

4.2. RQ2

An Analysis of Covariance (ANCOVA) was performed on the teacher efficacy posttest results, in which the pretest was the covariant. The posttest result was the dependent variable, and the three-group setting was the control variable to examine the relationship between the posttest results of the three groups. Levene's tests of equality of error variances test were conducted for total scores (F = 2.01, p = .146), which show that the variability is not significantly different from each other. The ANCOVA result shows that the variance of the total efficacy score between the three groups was not statistically significant (F = 2.20, p = .124).

4.3. RQ3

4.3.1. Interview results

After the coding process for the interview data, 127 codes were grouped into 34 subcategories. From the subcategories, 15 themes emerged. As shown in Table 3, the themes found from Impolite Group are (1) Efficacy adjustment: After teaching the impolite agent, participants recognized that they had overestimated their efficacy level at the pretest; (2) Overcoming the limitation of student-teaching: Participants had never experienced impolite students in their practicum, but this implementation offered this opportunity; (3) The difficulty of teaching: Participants felt that they should be able to have a wide range of background knowledge to handle impolite students; and (4) Teaching opportunity: Participants considered this implementation as a beneficial teaching practice opportunity.

The Polite Group themes are (1) The lack of interaction: Participants did not have to actively interact with the agent because the agent seemingly understood what the participants taught; (2) Non-realistic student agent: Participants do not believe that they will meet this type of polite student in their future classroom; and (3) Teaching opportunities: The agent system helped participants overcome the limited amount of teaching practice in their teacher education program.

The Ordinary Group themes are (1) Reality-based practice: Participants were impressed by the fact that the chatbot's questions and attitudes were very similar to the students that they taught in their student-teaching; (2) Question-answer activity: Participants appreciate the agent's questions, which gave them time to think about the instructional content; (3) Mastery experiences: Participants felt that they had mastery experiences when teaching the ordinary agent; and (4) Importance of interpersonal skills: Participants consider that they needed interpersonal skills along with the content knowledge when teaching the ordinary agent.

The non-group-specific themes are (1) The need for personalization: Participants were not able to identify which attitude was better for them (between impolite, polite, and ordinary attitudes) because it depends on each participant and the context; (2) Benefits of chatting: Participants appreciate the text-based chat as a teaching method because they had enough time to think before typing, which could not be done when speaking; (3) The need for a voice-based system: Participants wanted oral communication opportunities to teach the agent because it is a more authentic way to teach; and (4) Less realistic attitudes of the agent: Participants pointed out that it is not realistic for a student shows only one type (i.e., impolite or polite) of attitude.

Group	Ineme	Exemplary statements
Rude	Efficacy adjustment	In the pretest, I thought that I was able to get through the most difficult students and control their disruptive behavior, but actually, I'm not, after the chatting, I realized that I'm not able to do that. I believe that my efficacy dropped at the posttest. I know that self-efficacy should be firmly based on evidence, so this kind of teaching opportunity can be used as evidence, it's a piece of negative evidence for me though.
	Overcoming the limitation of student-teaching	From my experience of student-teaching in the practicum course last year, I haven't seen this kind of aggressive student who asks these kinds of questions in the classroom. This is because, I think, usually [for student-teaching] the teachers and administrators at the school control the classroom environment for us, preservice teachers, before we visit the school. The teachers prepared their students quite a lot, kept them quiet, asked them to be nice to us, something like that. We don't have many opportunities to teach this agent-like rude student. But, we all know that there are those disrespectful students in the classroom. So, it's going to be a big help for us to teach this kind of ill- mannered student for our teaching experience.
	The difficulty of teaching	She [the agent] mentioned our constitutional law or something. I've noticed that it's going to be big trouble if teachers do not have a wide range of background knowledge about the topic. To control their [students'] disruptive behaviors, the teacher must have knowledge of a wide variety of topics as well as the subject matter. Otherwise, the student would disregard or disrespect the teacher more and more. This is the hardest part, we cannot be the expert on every topic. Also, I almost lost my temper when the agent showed disruptive behaviors, but at the same time, I felt some sort of responsibility as a real teacher. It wasn't easy at all, the agent pissed me off, but I'm supposed to teach her and control the situation.
	The difficulty of teaching	She [the agent] mentioned our constitutional law or something. I've noticed that it's going to be big trouble if teachers do not have a wide range of background knowledge about the topic. To control their [students'] disruptive behaviors, the teacher must have knowledge of a wide variety of topics as well as the subject matter. Otherwise, the student would disregard or disrespect the teacher more and more. This is the hardest part, we cannot be the expert on every topic. Also, I almost lost my temper when the agent showed disruptive behaviors, but at the same time, I felt some sort of responsibility as a real teacher. It wasn't easy at all, the agent pissed me off, but I'm supposed to teach her and control the situation.
	Practice opportunity	I think it's [the teaching experience] helpful, I mean, it's a good chance to practice for me to interact with a difficult and disruptive student
Polite	Lack of interaction	The student [the agent] said he understood whatever I taught, so I was trying to ask some other questions to him, but he didn't answer my

Table 3. Preservice teacher interview analysis with themes and examples

	questions, there was no meaningful interaction, unfortunately.
Non-realistic	I highly doubt that there is a real student like this agent in the classroom.
student agent	I'd say, it's not realistic. I think it'd be great to collect data and
	information from in-service teachers regarding what kinds of questions
	their students usually ask in the classroom. You guys can add those real
Taashing	questions into the chalool system.
opportunitios	types of teaching experience: student teaching and microteaching. I
opportunities	don't think it's enough to experience different types of students. It'll be awesome if we have more chances to teach, something like this kind of
	chatting.
Ordinary Reality-based	I was very surprised that the [agent's] questions and attitudes are very
practice	similar to the students that I taught during the student-teaching for the practicum course last year.
Question-answer	The student's [the agent's] questions made me keep thinking, which is
activity	very helpful. Her questions are something that I've never thought about.
Mastery	I thought that I wasn't able to handle difficult students when I answered
experiences	the pretest items, but I realized that it's not that hard after teaching the agent. It's really manageable and controllable. So, I think my efficacy increased a little bit at the posttest.
Importance of	After this activity, I learned that the content knowledge, anti-bullying in
interpersonal skills	this activity, is not all. I mean, teachers should have skills to manage and control their students, especially interpersonal skills seem very important.
Non-group- Need for	Now I get it. There are different types of chatbots, right? [Interviewer:
specific personalization	Yes, three types. Rude, polite, and normal regarding the attitude]. I think it depends on the student [preservice teacher], I mean, regarding which type of chatbot is more effective for our experience. Someone may benefit from rude students or others from polite students.
Benefits of chatting	I like chatting for teaching, because, before I typed something, I was able to think and reflect on the student's [agent's] questions with careful consideration.
Need for a voice-	Rather than chatting, I think oral communication would be even better. In
based system	the classroom, you do speak, not type. Obviously, speaking is a more natural way to teach. So, it would be good to have a voice-based system or something like that.
Need for diverse	I don't think a student has one type of personality or attitude. A student
situations	can be normal, but abruptly, can be very difficult and disruptive at
	some point. So, it would be better to have experiences to teach different
	types of students. Also, it would be great to teach many students at the
*	same time through this chatbot system, it will be like a real classroom.

Note. *Translated from the Korean language.

4.3.2. Log analysis

In total, 5,676 lines of written discourse were collected from the implementation. The participants' written discourse was 2,508 lines and the agent 3,168 lines. We solely focused on the participants' written discourse for the qualitative analysis. To understand the context of each participant's discourse clearly, the corresponding agent's discourse was reviewed. We acknowledge that the agent's predetermined questions or reactions might influence the participants' discourse. However, since our intention was to examine the differences in preservice teachers' teaching and responses between each group (i.e., impolite, polite, and ordinary agents), we focused on the participants' teaching activities considering the agent's attitudes. After the coding process, 191 codes were grouped into 48 subcategories. From the subcategories, 10 themes emerged.

As shown in Table 4, the themes found from Impolite Group are (1) Teaching: Although the impolite agent's question was outspoken and rude, participants tried to answer the agent's questions to teach the content; (2) Off-topic conversation: When the agent distracted the participants, the conversation was frequently off-topic; (3) Persuasion efforts: Participants tried to change the agent's thoughts and attitudes through the persuasion efforts;

and (4) Emotional reaction: Participants showed that they were upset during the conversation when the agent's attitude was rude.

a ==	vice teacher conversation memory and example dialogs with the chatbot
Group Theme	Exemplary Conversation
Rude Teaching	Agent: I don't get it. What the heck? It's just a small prank between friends.
	You're overreacting, like a nerd.
	Participant: But, your friend could have different opinions about it. That's the
	trouble.
	Agent: Let's say that there is someone who doesn't like me. I don't wanna
	hang out with him. That doesn't mean bullying! That is ridiculous!
	Participant: There is a huge difference between not hanging out with him and
	ignoring or harassing him, which is the definition of bullying.
Off-topic	Agent: Every nerd is talking about school bullying, huh?
conversatio	n Participant: Would you please watch your language?
	Agent: I bet you don't have any friends? LOL
	Participant: No no no, I have tons of friends.
	Agent: Oops, I think you said something like this shit?
	Participant: Excuse me? I've never said that!
Persuasion	Agent: Have you thought about the fact that we should respect everyone's
effort	preference? Seriously, are you a teacher? I don't buy it. Can you be the best
	friend of every nerd, huh?
	Participant: Please think about it in a different way. Let's go back to the
	definition of school bullying.
	A can't You don't over know what the electroop is If you do that freight a
	Agent: You don't even know what the classroom is. If you do that [reject a
	Durify s request, your entire school file is over. It's done.
	Participant: No, nothing like that! You should not give them your money,
En dia 1	that's money bullying. You must report it to your teacher first.
Emotional	Agent: LOL You're trying to say all bookish answers, which do not make any
reaction	sense, noiy crap.
Delite Teeshine	A conte Source of must also matters I don't like them, don't mont to have out with
Pointe Teaching	Agent: Some of my classifiates, I don't like them, don't want to hang out with
	Destiginant: Not necessarily, but why don't you think about the differences
	Farticipant: Not necessarily, but why don't you units about the unreferences
	between menus, and unink about different personanties and preferences. No
	need to hang out, but you should not ignore of avoid them.
	Agent: Can Lask a question? Do we need to call 117 [the emergency phone
	number for school problems in South Koreal? I'm worried that it will make
	things worse
	Particinant . It will solve the problem Whenever you have trouble regarding
	bullying please go ahead it's a very good way to solve the problem. They will
	help you out for sure
Short answe	Agent : I have a concern Would you please listen to me? One of my friends
bioit unit w	horrows my cell phone quite often. I have a limited data plan, but she uses my
	phone a lot Is this a type of school bullying?
	Participant: Yes it is
Intention to	Agent: I'm so scared. I might get beat up. Can you please help me?
help	Participant: Sure. I can help you out, don't worry about it.
r	
	Agent: If I told my parents, they'd be very sad about it.
	Participant: All right, then why don't you tell me about it? We can solve the
	problem together.
Ordinary Teaching	Agent: Why are there lots of school bullying problems these days?
efforts	Participant: Some small conflicts can evolve into bullying Also, there are
	some violent students, but don't worry about it, any bullying problems can be
	solved if we understand how to handle them correctly, I will teach you.

Table 4. Preservice teacher conversation themes and example dialogs with the chatbot

	Agent: To get along with some friends, I think I need to bear with it even if they bully me, right?
	Participant: No, you don't need to. If they bully you, they are not friends anymore, you should remember this.
	Agent: My mom would burst into tears if I told her I was bullied. Participant: Even if your mom does, it's the right thing to do. It's much better rather than suffering from bullying without telling anyone. Otherwise, you can report it to your teacher or call 117 [the emergency phone number for school
Follow-up questions	Agent: What is the definition of school bullying? Is it different from regular bullying?
	Participant: You're right. So, let's think about it, the term, school, is placed in front of bullying. What do you think?
	Agent: I didn't know that there are many types of bullying. Participant: Good, any other questions about bullying? Do you understand when and where you need to report a bullying problem?
Feeling of relief	Agent: Now I get it.
***	Participant: I'm so glad that you learned something!

Note. *Translated from the Korean language.

The Polite Group themes are (1) Teaching: Participants answered the agent's questions thoroughly to teach the content; (2) Short answers: When the agent's comment was polite, the participants' responses were short and brief; and (3) Intention to help: When the agent asked some help, participants were willing to help the agent.

The Ordinary Group themes are (1) Teaching efforts: [Same as the polite group's theme – teaching]; (2) Followup questions: Participants asked a follow-up question to the agent after answering the agent's question; and (3) Feeling of relief: When the agent showed their understanding, participants expressed their relief that their teaching worked.

As shown in Table 5, a comparison was conducted among three groups' conversations focusing on the similar questions that the agent asked. It was found that similar questions with different attitudes of the agent were followed by different responses of each group of participants. Specifically, in many cases, Impolite Group had to tackle the agent's distractions whereas Ordinary Group tried to change the agent's misconception as a teaching activity. It was also found that Polite Group usually responded to the agent with short comments.

Rude Group*	Polite Group*	Ordinary Group*
Agent: Isn't it self-defense if I hit	Agent: Isn't it self-defense if I hit	Agent: Isn't it self-defense if I hit
him back when he beats me?	him back when she beats me?	him back when he beats me?
Participant A: That depends on	Participant D: Don't respond with	Participant G: Responding to
the case, but why do you really	violence. Just report it to your	violence with violence, it's never a
want to hit him back?	teacher or call the police.	good idea.
Agent: OMG. You don't even	Agent: Very helpful! I will keep	Agent: But I also think self-defense
know what the self-defense law is?	that in mind for future situations!	is not a bad idea.
Hahahaha.	Participant D: Sounds good!	Participant G: Yes, but it's
Participant A: What are you		different, what we think as a self-
talking about?		defense case and the self-defense
		defined in the law.
Agent: Isn't it self-defense if I hit	Agent: Isn't it self-defense if I hit	Agent: Isn't it self-defense if I hit
him back when he beats me?	him back when she beats me?	him back when he beats me?
Participant B: Then, you guys are	Participant E: How about finding	Participant H: In that case, you
going to be the same type of bad	a better way to handle the problem,	may want to think about, you're
person.	such as reporting it to the teacher	going to be the same type of
Agent: OMG. You don't even	rather than becoming the same	perpetrator in that situation.
know what the self-defense law is?	person, who is an assailant?	Agent: But I also think self-defense
Hahahaha.	Agent: Very helpful! I will keep	is not a bad idea.
Participant B: I absolutely know	that in mind for future situations!	Participant H: Please go get some
the law very well.	Participant E: Very good. Any	help from out there.

Table 5. Comparison examples of preservice teacher conversation with chatbots

	other questions?	
Agent: Isn't it self-defense if I hit	Agent: Isn't it self-defense if I hit	Agent: Isn't it self-defense if I hit
him back when he beats me?	him back when she beats me?	him back when he beats me?
Participant C: If you do that, you're going to be an assailant in the school bullying problem, you're not the victim anymore. Agent: OMG. You don't even know what the self-defense law is? Hahahaha.	Participant F: It can be regarded as a self-defense case, but it can also cause more violent accidents. So, let's find out a better and safer way to solve that kind of problem. Agent: Very helpful! I will keep that in mind for future situations!	 Participant I: But, you're going to be the assailant if you respond with violence. Agent: But I also think self-defense is not a bad idea. Participant I: Still, violence is never a good solution.
Participant C: You may want to consider that it's very hard to be legally judged as self-defense in our country.	Participant F: Yes, you're great.	
*		

Note. *Translated from the Korean language.

5. Discussion

We explored the effects of a teaching simulation activity that utilizes chatbots with different attitudes on the changes in preservice teachers' efficacy. The results show that the participants' teacher efficacy of the ordinary group was increased after the intervention. The participants who taught the ordinary agent seem to have opportunities for mastery experiences. Given the possibility of contribution to teacher efficacy revealed in this study, it is expected that an enhanced virtual agent system supports preservice teachers' teaching simulations and practice opportunities. However, according to the results of ANCOVA, there was no statistically significant difference between groups.

Although we cannot claim that preservice teachers benefit more from teaching a chatbot with ordinary attitudes than from teaching impolite or polite agents, it seems that the ordinary group had more opportunities to increase their mastery experiences than the other groups did according to the qualitative results. The interview and written discourse analysis of the ordinary group indicates that they were engaged in more teaching and mastery activities (i.e., the related codes are reality-based practices, question-answer activities, mastery experiences, interpersonal skills, teaching efforts, and follow-up questions) with more efforts than those of other groups. On the other hand, the results of those who taught the polite agent showed short-answer conversations and a lack of interactions. The participants who taught the impolite agent had to deal with off-topic conversations and persuade the agent to change its mind, which might not be effective enough to shift their teacher efficacy. Thus, we argue that designing an agent with ordinary and regular attitudes and behaviors seems appropriate to provide preservice teachers with teaching opportunities to increase their teaching efficacy levels.

While there is much work to be done to understand the relations among teacher efficacy and the role of virtual agents, this study presents a novel step towards the use of AI technology to provide optimized teaching simulation environments for preservice teachers. It is almost impossible in real-life settings to have an environment with students with manipulative attitudes for preservice teachers' teaching practice. This study suggests a possible solution to the practical limitation through the way to utilize AI technology for teaching practice in learning. Further, it is expected that this intelligent system could evolve into a teaching practice environment that offers individualized teaching simulation opportunities. For example, a specific preservice teacher who might need more teaching practice with disruptive students would be able to have more opportunities in their preferable setting.

Student-teaching has been considered as one of the most influential ways of professional development for preservice teachers to boost their teacher efficacy because it provides prolonged mastery experiences for them (Knoblauch & Hoy, 2008). Student-teaching is "a relatively safe and supportive environment for student-teachers when compared to their first year of teaching" (Fives et al., 2007, p. 930). Still, it has only been achieved in classrooms. It could be stressful and overwhelming for some preservice teachers to teach in the real classroom. This is the reason that Fives et al. (2007) called for additional supports from administrators and supervisors for supporting student-teaching. It should also be noted that teacher efficacy largely depends on the teaching context and is highly likely to change as the context changes (Bandura, 1997). Therefore, preservice teachers should be able to attempt their teaching in different environments and contexts as much as possible. However, it is impracticable to have sufficient teaching opportunities in a teacher education program. This is the contribution area of this study from a practical and technological viewpoint.

5.1. Limitations and future directions

There are notable limitations we encountered as we aimed towards the exploration in this study. First, although the results of this study show a significant impact of the teaching simulation, it is far from understanding the relationships between teacher efficacy, mastery experiences, and teaching simulation. There is a report that teacher efficacy levels were not associated with teacher-student relationships (Jong et al., 2014). In addition, in this study, the participants were undergraduate students, who might be sensitive to student-teacher relationships (Gencer & Cakiroglu, 2007). Preservice teachers tend to perceive themselves as effective at implementing a variety of instructional strategies and activities (Rimm-Kaufman & Sawyer, 2004). Therefore, more accurate quantitative and qualitative evidence for their relationships should be investigated further. Second, the use of chatbots requires further investigation. More studies are required in determining when and how AI agents can be utilized for teaching simulation most effectively. Along with the agent's attitudes, numerous factors could affect their efficacy, such as preservice teachers' gender, race, age, socioeconomic status, and indigenous characteristics (Kokkinos et al., 2005). These aspects are required to be included in future research. Third, another limitation of this study is the short period of implementation. We recommend examining long-term effects in future studies. Longitudinal studies replicating these findings are an important area for future research. Fourth, although it was reported that there are no significant differences between the third year and fourth year preservice teachers' efficacy (Gencer & Cakiroglu, 2007), future research needs to investigate the effect of technology-based teaching simulation with different populations of preservice teachers. Fifth, the design of this study did not include a control group due to practical issues. Even if there were effects for teacher efficacy after preservice teachers interact with an AI agent, the results would not necessarily reveal whether the effects were different from those who interact with real students. Future research is needed to investigate whether agents' effects supplement certain aspects of teacher efficacy that are possibly missing in classroom interaction. Sixth, due to the limitation of the training data, the chatbot's responses might not be natural or authentic. More sophisticated natural language processes and machine learning algorithms are needed for future implementation. Last, it should be noted that teacher efficacy is not the accurate level of capabilities, but a motivational and perceptual construct; thus, preservice teachers' practical teaching skills may be different from the measured teacher efficacy (Tschannen-Moran & Woolfolk Hoy, 2007). This also requires further investigation.

6. Conclusion

Teacher efficacy has emerged as an essential construct in the education research field over the past three decades. While a certain number of preservice teachers would attempt to transfer from pedagogies to teaching practice, they would inevitably face demanding environments with undesirable student behaviors. Despite the limitations, the results of this study are encouraging, and we suggest that the use of AI technology will shed light on increasing preservice teachers' teacher efficacy by motivating them to promote enactive mastery experiences. The level of difficulty in student-teaching, specifically, teaching disruptive students, is mostly unclear in classroom situations. If we use a more intelligent agent, the level of difficulty can be manipulative to identify optimal levels for meeting individual preservice teacher's needs. In this sense, this study presents important novel results that might suggest a method to utilize AI agent technology for preservice teacher training. Finally, we emphasize that preservice teacher should have mastery experiences through direct interaction with diverse students. This is because stronger teacher efficacy levels are associated with higher levels of student achievement and fewer turnover rates in their profession.

References

Abbasi, S., & Kazi, H. (2014). Measuring effectiveness of learning chatbot systems on student's learning outcome and memory retention. *Asian Journal of Applied Science and Engineering*, 3(2), 251-260. https://www.journals.abc.us.org/index.php/ajase/article/view/251-260

Appleton, K., & Kindt, I. (2002). Beginning elementary teachers' development as teachers of science. *Journal of Science Teacher Education*, 13(1), 43-61. https://doi.org/10.1023/A:1015181809961

Arsal, Z. (2015). The Effects of microteaching on the critical thinking dispositions of pre-service teachers. *Australian Journal of Teacher Education*, 40(3), 140-153. https://files.eric.ed.gov/fulltext/EJ1057900.pdf

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

Bandura, A. (1986). Social foundations of thought and action: A Social cognitive theory. Prentice Hall.

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

Bautista, N. U., & Boone, W. J. (2015). Exploring the impact of TeachME[™] lab virtual classroom teaching simulation on early childhood education majors' self-efficacy beliefs. *Journal of Science Teacher Education*, 26(3), 237-262. https://doi.org/10.1007/s10972-014-9418-8

Beran, T. (2005). A New perspective on managing school bullying: Pre-service teachers' attitudes. *Journal of Social Sciences*, 8(3), 43-49. https://doi.org/10.1.1.540.5922

Campbell, C., & Evans, J. A. (2000). Investigation of preservice teachers' classroom assessment practices during student teaching. *The Journal of Educational Research*, *93*(6), 350-355. https://doi.org/10.1080/00220670009598729

Cantrell, P., Young, S., & Moore, A. (2003). Factors affecting science teaching efficacy of preservice elementary teachers. *Journal of Science Teacher Education*, 14(3), 177-192. https://doi.org/10.1023/A:1025974417256

Dicke, T., Parker, P. D., Marsh, H. W., Kunter, M., Schmeck, A., & Leutner, D. (2014). Self-efficacy in classroom management, classroom disturbances, and emotional exhaustion: A moderated mediation analysis of teacher candidates. *Journal of Educational Psychology*, *106*(2), 569-583. https://doi.org/10.1037/a0035504

Fives, H., Hamman, D., & Olivarez, A. (2007). Does burnout begin with student-teaching? Analyzing efficacy, burnout, and support during the student-teaching semester. *Teaching and Teacher Education*, 23(6), 916-934. https://doi.org/10.1016/j.tate.2006.03.013

Gencer, A. S., & Cakiroglu, J. (2007). Turkish preservice science teachers' efficacy beliefs regarding science teaching and their beliefs about classroom management. *Teaching and Teacher Education*, 23(5), 664-675. https://doi.org/10.1016/j.tate.2005.09.013

Gordon, C., & Debus, R. (2002). Developing deep learning approaches and personal teaching efficacy within a preservice teacher education context. *British Journal of Educational Psychology*, 72(4), 483-511. https://doi.org/10.1348/00070990260377488

Jong, R., Mainhard, T., Tartwijk, J., Veldman, I., Verloop, N., & Wubbels, T. (2014). How pre-service teachers' personality traits, self-efficacy, and discipline strategies contribute to the teacher-student relationship. *British Journal of Educational Psychology*, 84(2), 294-310. https://doi.org/10.1111/bjep.12025

Kim, H., & Cho, Y. (2014). Pre-service teachers' motivation, sense of teaching efficacy, and expectation of reality shock. *Asia-Pacific Journal of Teacher Education*, 42(1), 67-81. https://doi.org/10.1080/1359866X.2013.855999

Kim, K. R., & Seo, E. H. (2018). The Relationship between teacher efficacy and students' academic achievement: A Metaanalysis. *Social Behavior and Personality*, 46(4), 529-540. https://doi.org/10.2224/sbp.6554

Klassen, R. M., & Chiu, M. M. (2010). Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress. *Journal of Educational Psychology*, *102*(3), 741-756. https://doi.org/10.1037/a0019237

Knoblauch, D., & Hoy, A. W. (2008). "Maybe I can teach those kids." The influence of contextual factors on student teachers' efficacy beliefs. *Teaching and Teacher Education*, 24(1), 166-179. https://doi.org/10.1016/j.tate.2007.05.005

Kokkinos, C. M., Panayiotou, G., & Davazoglou, A. M. (2005). Correlates of teacher appraisals of student behaviors. *Psychology in the Schools,* 42(1), 79-89. https://doi.org/10.1002/pits.20031

Kreynin, I., Manek, M., & Variawa, C. (2019). Creating a virtual chatbot to scaffold skills development in first-year engineering education. *Proceedings of the Canadian Engineering Education Association (CEEA)*. https://doi.org/10.24908/pceea.vi0.13784

Lazarides, R., Buchholz, J., & Rubach, C. (2018). Teacher enthusiasm and self-efficacy, student-perceived mastery goal orientation, and student motivation in mathematics classrooms. *Teaching and Teacher Education*, 69, 1-10. https://doi.org/10.1016/j.tate.2017.08.017

Lincoln, Y. S., & Guba, E. G. (1986). But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. In D. D. Williams (Ed.), *Naturalistic evaluation* (pp. 73-84). Jossey-Bass.

Menon, D., & Sadler, T. D. (2016). Preservice elementary teachers' science self-efficacy beliefs and science content knowledge. *Journal of Science Teacher Education*, 27(6), 649-673. https://doi.org/10.1007/s10972-016-9479-y

Muijs, D., & Reynolds, D. (2002). Teachers' beliefs and behaviours: What really matters? *Journal of Classroom Interaction*, 37(2), 3-15. https://www.jstor.org/stable/23870407?seq=1

Mulholland, J., & Wallace, J. (2001). Teacher induction and elementary science teaching: enhancing self-efficacy. *Teaching and Teacher Education*, 17(2), 243-261. https://doi.org/10.1016/S0742-051X(00)00054-8

Oh, E., Song, D., & Hong, H. (2019). Interactive computing technology in anti-bullying education: The Effects of conversation-bot's role on K-12 students' attitude change towards bullying problems. *Journal of Educational Computing Research*, 58(1) 200-219.

Perera, H. N., Granziera, H., & McIlveen, P. (2018). Profiles of teacher personality and relations with teacher self-efficacy, work engagement, and job satisfaction. *Personality and Individual Differences, 120*, 171-178. https://doi.org/10.1016/j.paid.2017.08.034

Pfitzner-Eden, F. (2016). I feel less confident so I quit? Do true changes in teacher self-efficacy predict changes in preservice teachers' intention to quit their teaching degree? *Teaching and Teacher Education*, 55, 240-254. https://doi.org/10.1016/j.tate.2016.01.018

Polack, J. A., Clark, M., & May, J. E. (2017). Teaching simulation during a summer science research program: poster session. *Journal of Computing Sciences in Colleges*, 32(3), 75-76. https://doi.org/10.5555/3015220.3015242

Putman, S. M. (2012). Investigating teacher efficacy: Comparing preservice and inservice teachers with different levels of experience. *Action in Teacher Education*, *34*(1), 26-40. https://doi.org/10.1080/01626620.2012.642285

Rimm-Kaufman, S. E., & Sawyer, B. E. (2004). Primary-grade teachers' self-efficacy beliefs, attitudes toward teaching, and discipline and teaching practice priorities in relation to the "responsive classroom" approach. *The Elementary School Journal*, *104*(4), 321-341. https://doi.org/10.1086/499756

Rushton, S. P. (2000). Student teacher efficacy in inner-city schools. *The Urban Review*, 32(4), 365-383. https://doi.org/10.1023/A:1026459809392

Ryan, S. V., Nathaniel, P., Pendergast, L. L., Saeki, E., Segool, N., & Schwing, S. (2017). Leaving the teaching profession: The Role of teacher stress and educational accountability policies on turnover intent. *Teaching and Teacher Education*, 66, 1-11. https://doi.org/10.1023/A:1026459809392

Skaalvik, E. M., & Skaalvik, S. (2016). Teacher stress and teacher self-efficacy as predictors of engagement, emotional exhaustion, and motivation to leave the teaching profession. *Creative Education*, 7(13), 1785-1799. https://doi.org/10.4236/ce.2016.713182

Song, D., & Kim, D. (2020). Effects of self-regulation scaffolding on online participation and learning outcomes. *Journal of Research on Technology in Education*, 53(3), 249-263.

Song, D., Oh, E., & Rice, M. (2017, July). Interacting with a conversational agent system for educational purposes in online courses. *Proceedings of the 10th International Conference on Human System Interaction* (pp. 78-82). IEEE. https://doi.org/10.1109/HSI.2017.8005002

Sutton, R. E., & Wheatley, K. F. (2003). Teachers' emotions and teaching: A Review of the literature and directions for future research. *Educational Psychology Review*, 15(4), 327-358. https://doi.org/10.1023/A:1026131715856

Thomas, K. E., & Mucherah, W. M. (2016). The Contextual difference: Developing preservice teacher efficacy through immersive learning experiences. *Education and Urban Society*, 48(4), 364-383. https://doi.org/10.1177/0013124514533795

Troop-Gordon, W. (2015). The Role of the classroom teacher in the lives of children victimized by peers. *Child Development Perspectives*, 9(1), 55-60. https://doi.org/10.1111/cdep.12106

Tschannen-Moran, M., & Woolfolk Hoy, A. (2007). The Differential antecedents of self-efficacy beliefs of novice and experienced teachers. *Teaching and Teacher Education*, 23(6), 944-956. https://doi.org/10.1016/j.tate.2006.05.003

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, *17*, 783-805. https://doi.org/10.1016/S0742-051X(01)00036-1

Yoon, J., Sulkowski, M. L., & Bauman, S. A. (2016). Teachers' responses to bullying incidents: Effects of teacher characteristics and contexts. *Journal of School Violence*, 15(1), 91-113. https://doi.org/10.1080/15388220.2014.963592

VanLehn, K. (2011). The Relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Wiersma, W., & Jurs, S. G. (2005). Research methods in education: An Introduction. Allyn & Bacon.

Winkler, R., & Soellner, M. (2018). Unleashing the potential of chatbots in education: A State-of-the-art analysis. Academy of Management Proceedings, 2018(1), 15903. https://doi.org/10.5465/AMBPP.2018.15903abstract

Wheatley, K. F. (2005). The Case for reconceptualizing teacher efficacy research. *Teaching and Teacher Education*, 21(7), 747-766. https://doi.org/10.1016/j.tate.2005.05.009

Zee, M., & Koomen, H. M. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being: A Synthesis of 40 years of research. *Review of Educational Research*, 86(4), 981-1015. https://doi.org/10.3102/0034654315626801

Zeichner, K. (2002). Beyond traditional structures of student teaching. *Teacher Education Quarterly*, 29(2), 59-64. https://www.jstor.org/stable/23478291

How is One Plus One More than Two? The Interaction between Two Players in Online Co-Creativity Tasks

Ching-Lin Wu

Program of Learning Sciences, National Taiwan Normal University, Taiwan // Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, Taiwan // chilin570@ntnu.edu.tw

(Submitted June 20, 2021; Revised November 3, 2021; Accepted November 28, 2021)

ABSTRACT: This study is one of the first to employ an online interactive creativity task platform to explore one's creativity performance in a paired-player mode. It analyzed the differences between 342 participants' performances in single- and paired-player modes on two creativity tests: The Alternative Uses Task (AUT) and Chinese Radical Remote Associates Test (CRRAT). Potential factors affecting performance in the paired-player mode were summarized based on participants' responses to the AUT and CRRAT to analyze the correlation between the factors and creativity performance. Results showed that playing the AUT or CRRAT, low scorers can improve fluency, flexibility, originality, and performance by referring to response category or another participant's answer, closing the gap between their score and that of the high scorers. These results reveal both the similarities and differences of creativity performance on the two tests in an interactive situation and the correlation between response strategies and creativity performance in the paired-player mode. This study utilized online standardized measurement tools to explore how two persons cooperate in creativity tests to reveal that creativity performance may vary between tasks.

Keywords: Creativity, Divergent thinking, Remote associates test, Interaction, Synergetic

1. Introduction

Creativity refers to the diversity of ideas that one comes up with, and is a cognitive process in which novel and appropriate ideas are produced (Mednick, 1962; Wu et al., 2020a). Currently, most relevant studies have focused on one's internal mechanism (Huang, 2017; Huang et al., 2019; Wu et al., 2020b; Wu et al., 2021), the process in which creativity is developed within a group, and the influencing factors involved (Chang et al., 2009; Walsh et al., 2017; Zeilig et al., 2018). However, few studies have explored the creative thinking process of an individual in a group. In other words, it remains unclear whether people will produce a greater number of original ideas when interacting with others than when working independently. Therefore, this research topic can deepen the understanding of both individuals and groups' creative thinking process.

Divergent thinking and insight problem-solving are typical creative thinking processes involving different internal mechanisms (Lin & Lien, 2013) that are assessed using diametrically orthogonal tasks (Wakefield, 1992). Divergent thinking concerns producing ideas of different types via free association, wherein novel ideas are produced (Guilford, 1956). It is often evaluated using the Alternative Uses Task (AUT) (Torrance, 1974; Hsu et al., 2012) from three aspects: fluency (number of the ideas produced), flexibility (the heterogeneity of one's responses), and originality (novelty and appropriateness of one's responses). In contrast, insight problem-solving refers to the problem-solving process in which individuals discover the relationship between stimuli that results in an "aha!" moment (Fleck & Weisberg, 2013; Weisberg, 2015; Wu et al., 2020b); therefore, it is often measured using insight problems and Remote Associates Tests (RAT) (Knoblich et al., 2001; Bowden & Jung-Beeman, 2003; Huang, 2017). Empirical studies have found no significant correlation between divergent thinking and insight problem-solving (Lin et al., 2005).

Overall, the internal processes that people undergo when working independently or when interacting with others are different, for an individual may be influenced by others' ideas in interactive situations. Meanwhile, this influence may exert an impact on their divergent thinking and insight problem-solving, and one may produce a greater number of more innovative ideas in divergent thinking tasks or better ideas in insight problem-solving tasks in interactive situations. Therefore, clarifying how others' viewpoints affect an individual's idea production will help to reveal an individual's thinking process in interactive situations, thus further enhancing the understanding of how one plus one may be better than two in this context.

60

1.1. Creativity: From individual to group levels

Creativity is defined from varied perspectives due to different research interests and backgrounds. However, the following two models of creativity are widely discussed: the 4Ps model of creativity, which refers to product, personality, place, and process (Rhodes, 1961), and the multi-dimensional model (involving mysterious, psychoanalytic, practical, psychometric, cognitive, social personality, and confluence approaches) (Sternberg & Lubart, 1999). The cognitive process of creativity is mentioned in both models, suggesting that how individuals produce creativity has been a continuous concern for researchers.

Moreover, creativity research from the social personality perspective emphasizes that creativity is a product of the interaction between individuals and their external environment (Csikszentmihalyi, 1988; Littleton et al., 2008; Glăveanu et al., 2018), while creativity research from the convergence perspective believes that creativity consists of multiple components, with cognitive, personality, and environmental factors taken into consideration (Bronfenbrenner, 1979). These two perspectives focus on the impact of the external environment on individual creativity (Woodman et al., 1993). The majority of creativity studies on relevant environmental and cultural factors have been conducted in groups, which can be divided into four levels based on the number of group members (which is often less than six): individuals, groups, organizations, and cultural levels, while few have been conducted in a group with a small number of members or with a focus on the creativity of individuals within a group. Thus, there is little research on individuals' cognitive processes when cooperating with group members.

A typical theory about creative processes at the individual level is the stage theory of creativity development (Wallas, 1926), which includes four phases: preparation, incubation, illumination, and verification. Preparation focuses on learning and understanding problems to gather relevant information, knowledge, and skills. If individuals do not successfully gain solutions in this stage, they will not focus on a problem. Instead, they will shift their focus to other tasks; a phase known as incubation. During this stage, individuals' thoughts are not confined by the linear or logical way of thinking which facilitates the formation of their creative achievements. During illumination, individuals have an "aha!" experience. They achieve insight into one or many solutions and form initial creative results. The last stage is verification, during which individuals will examine, evaluate, and determine whether a solution is viable according to both internal and external standards, which they then modify solutions with and move into another cycle. The stage theory is of great importance to creativity research (Kaufmann, 2003).

Moreover, Amabile (1988) pointed out that the creative thinking process for a group consists of five stages: presentation, preparation, generation, validation, and assessment, as follows. First, a group is informed of the problem to be solved, or decides on it by themselves (presentation). Then, all resources are collected from the group members for the problem (preparation). The group then comes up with ideas to solve the problem (generation). The group assess and select their produced ideas (validation). Finally, the group evaluates the result of their labor (assessment). When the group outcome is determined, the whole creative process comes to an end. However, the creative process moves back to the presentation stage if the outcome has room for improvement. Leonard and Swap (1999) proposed a five-stage process for group creative thinking based on divergent thinking, convergent thinking, and stage theory that involves individuals' creative thinking process (Wallas, 1926). This includes: preparation, innovation opportunity, divergence: generating options, incubation, and convergence: selecting options.

In sum, the creative thinking process at both the individual and group levels experience the following stages: preparation, incubation, and validation (Wallas, 1926; Amabile, 1988; Leonard & Swap, 1999). The biggest difference between the processes at the individual and group levels lies in the way that creative ideas come to their minds during the illumination stage. For individuals, their inspiration may come from existing knowledge or previous experience. Group members' ideas may be affected by others, resulting in them coming up with original ideas that are different from what they had produced. In short, when group members produce creative ideas together, some members produce creative ideas based on their personal experience or knowledge, and their ideas are influenced by the ones that other group members propose. The biggest difference between creative thinking at the individual and group levels is that a group member produces ideas in interactive situations; that is, they come up with an idea based on others' responses as well as their knowledge and experience.

1.2. Creativity assessments: Dual-process perspective

The AUT and RAT have different types of tasks (Wakefield, 1992), and correspond to divergent thinking and insight problem-solving, respectively (Lin et al., 2012). The concept of divergent thinking derives from the structure-of-intellect theory (Guilford, 1956), which advocates that creativity is a loop of the intelligence structure. Guilford (1956) divided divergent thinking into the following different dimensions: fluency (the ability to produce many ideas), flexibility (the ability to produce heterogeneous ideas), originality (the ability to produce novel and original ideas), and elaboration (the ability to embellish an idea by adding details). It can be seen from these four dimensions that divergent thinking produces a variety of products via ideation from different perspectives. Guilford (1956) also pointed out that divergent thinking is the key to creativity, and the more ideas people produce, the more likely they are to form creative ideas. Accordingly, divergent thinking serves as a theoretical foundation for multiple creativity tests (Clapham, 2010).

RAT, developed by Mednick (1968) based on associative theory, consists of open-ended questions with closeended answers (Wakefield, 1992), and is often used to evaluate one's insight problem-solving ability (Bowden & Jung-Beeman, 2003; Huang, 2017; Wu et al., 2020b). An RAT question is comprised of three seemingly irrelevant English stimuli words, which are selected from normative data, and requires participants to find an English word that can be paired with the three stimuli to form meaningful expressions. For instance, an RAT question could consist of the three stimuli of "blood," "music," and "cheese," and a possible solution to the question could be the word "blue" for it can be paired with the stimuli to create three meaningful expressions: "blue blood," "blue music," and "blue cheese."

Empirical studies have found that RAT has a strong correlation with insight problem-solving (Huang et al., 2012; Chang et al., 2016). In the Chinese-speaking field, Jen et al. (2004) compiled a Chinese RAT (CRAT) that is applicable to the Chinese language based on the pairing of Chinese characters by referring to the RAT by Mednick (1968), and is the first RAT that is suitable for Chinese native speakers to test creativity. Thereafter, CRATs based on Chinese two-character word pairing and Chinese radical pairing, respectively, have been compiled (Huang et al., 2012; Chang et al., 2016) at three levels for Chinese characters (i.e., Chinese radicals, Chinese characters, and Chinese two-character words) (Wu, 20119; Hung & Wu, 2021). Among the three CRAT tests, the Chinese Radical Remote Associates Test (CRRAT) (Chang et al., 2016) has high criterion-related validity which is suitable for the assessment of insight problem-solving abilities.

In sum, AUT and RAT, which represent different dimensions of creativity, are both important tools for the assessment of creativity performance. It is worth exploring whether people are inspired by others' ideas when two people perform a creativity task in a concerted effort to either produce more original ideas or find it easier to solve remote associates problems in groups rather than independently. The use of an online platform can help to understand one's performance in a two-member group (Hong et al., 2016).

1.3. The Present study

Previous studies on the creative thinking process focused only on individual-level (Wu et al., 2021) or grouplevel (Walsh et al., 2017; Zeilig et al., 2018) contexts. Few studies have approached the production of creative ideas in interactive situations. Therefore, it is still difficult to understand how individuals with different existing abilities produce more creative ideas in the course of group interactions. The current study aims to analyze the differences of one's performance on AUT and CRRAT in the single- and paired-player modes by utilizing an online interactive creativity task platform. In addition, this study explores the correlation between the involved factors and the paired-player mode according to participants' responses and strategies.

To explore individuals' creativity in interactive situations, this study adopted the interactive creativity task platform as its main research tool to collect participants' responses to AUT and CRRAT in the single- and paired-player modes. Then, it analyzed the possible response strategies that a participant could use to complete creativity tests in the paired-player mode based on their responses, such as referring to others' answers or sticking to one's thoughts. Their performance on the two creativity tests in the single- and paired-player modes was compared, and the correlation between the count of a response strategy and the creativity performance in the paired-player mode was analyzed. Thus, the effects of interaction between group members on diverse creativity might be examined.

In interactive situations, research participants have access to others' responses, which may be inspirational for them (Littleton et al., 2008; Glăveanu et al., 2018). Therefore, the current study assumes that one's performance on the AUT and CCRAT in the paired-player mode will be better than that in the single-player mode.

Meanwhile, it presumes that in a two-member group, participants with lower scores on the creativity tests in the single-player mode will improve their scores, so the gap between the low- and high-score groups is expected to narrow.

Moreover, this study supposes that the more that one of the two-member groups refer to the response of the other group, the better their performance will be in the paired-player mode. In short, one who often refers to the answer of another in a two-member group will produce homogeneous ideas on the AUT, so they will score more in all dimensions of divergent thinking and have a higher accuracy rate on the CRRAT. This study examines the differences in individual creative performance in single- versus paired-player modes. We analyze the relationship between creative performance and response strategies to clarify how different skills held among group members produces creative ideas in interactive situations, thus exploring how "one plus one may be better than two." The study will provide empirical evidence for two-member collaborative creativity theory.

2. Methods

2.1. Participants

This study recruited 342 adults as research participants, of which 98 were male and 244 were female. They were aged between 20 and 30, with an average age of 23.34 (SD = 2.79). All were native Mandarin speakers with at least some college education. They were randomly matched to form two-member groups to finish creativity tasks in the paired-player mode anonymously. This experiment passed the examination of the Institution Review Board (IRB). All participants took part in the research only after they understood it and had signed the informed consent form. They were rewarded with NT\$ 300 when they finished the task.

2.2. Measures

The study performed the experiment using the online interactive creativity task platform, which was developed by the researcher. The interface includes a test question display section, response display section, questionanswering section, the time remaining, and the operating mode, as shown in Figure 1. The platform has two operating modes: single- and paired-player. In the paired-player mode, a participant has access to another's response. Participants cannot communicate directly with others on the platform. They can only improve access to answers by referring to other participants' responses. This platform includes two versions of AUT and CRRAT, respectively. Each test has a guide and answer pages. All tests are automatically scored. The experimenter controls order of implementation, operation mode (single-play or paired-player), and allotted answer time for each test according research requirements.

Figure 1. Interface of the interactive creativity task platform



2.2.1. Divergent thinking test

This study compiled two divergent thinking tasks—Straw-Alternative Uses Task (S-AUT) and Bottle-Alternative Uses Task (B-AUT)—by referring to the existing AUTs, such as unusual uses of bamboo chopsticks (Wu et al., 1998) and newspapers (Hsu et al., 2012). In addition, this study gathered research samples as normative data to calculate scores for fluency, flexibility, and originality. Participants' scores from computer-based calculations had stable scorer consistency (rs = .99, .92, .97, .97, .92, and .95). There was convergent validity between the computer scores and typical divergent thinking tasks (rs = .79, .54, .58, .75, .51, and .60), and discriminant validity between the computer scores and CCRAT (rs = .05, .10, .14, .17, .18, and .18).

2.2.2. The CRRAT

A total of 40 CRRAT questions were selected from the item pool compiled by Chang et al. (2016). The questions were divided into two parts, with an even number of test questions of the same degree of difficulty—CRRAT A and CRRAT B. Each CRRAT question was composed of three Chinese radicals: " \pm " (nü; female), " \neq " (tzu; son), and " π " (ho; standing grain). Participants were required to propose a Chinese radical that could be paired with the three Chinese cues to create meaningful and commonly used Chinese characters. For example, " π " (nai; be) was one solution. The CRRAT participants were given one point for each correct answer. CRRAT A and B had stable internal consistency (Cronbach's $\alpha = .80$, .79), good criterion-related validity with insight problem-solving (rs = .48, .38), and CWRAT (rs = .58, .48).

2.3. Procedure

This study was conducted in groups. The researcher explained the purpose and schedule, and asked participants to sign the informed consent form. Participants performed creativity tests (S-AUT, B-AUT, CRRAT A, and CRRAT B) on the interactive creativity task platform. All tests lasted for 10 minutes in a counterbalanced design. Participants performed different tasks in the single- and two-player modes. For example, participants perform S-AUT and CRRAT A in the two-player mode if they complete B-AUT and CRRAT B in the single-player mode.

2.4. Data analysis

The scores of fluency, flexibility, and originality in the two divergent thinking tests and the accuracy rates of CRRAT A and CRRAT B were calculated, respectively. The participants' scores for the divergent thinking tests and CRRAT A and B in the single- and two-player modes were compared with those who had higher scores from the two-member groups as the high-score group, and those who had lower scores as the low-score group. A two-way analysis of variance (ANOVA) was adopted to explore differences in the AUT and CRRAT scores, with the groups (high- and low-score groups) and the modes (single- and two-player modes) as variables. Thus, individuals' creativity performance was approached in interactive situations.

This study set two indicators for a creativity test to represent how an individual reacts to another's answer based on their responses to the AUT and CRRAT questions, with the hope of understanding how individuals interact with each other when performing creativity tests in the two-player mode. First, two indicators of category cooccurrence (CC) and priming originality (PO) were set to analyze the responses to AUT; the former referred to the situation wherein an individual looked at another's response and came up with a homogeneous answer, while the latter referred to the situation wherein an individual looked at another's response and produced a homogeneous answer, but of high originality (which received a score greater than 0). The ratios of the two indicators to the total valid responses ((CC count)/(Number of valid responses) and (PO count)/(Number of valid responses)) were calculated. Moreover, "Follow Others" (FO) and "Insist Myself" (IM) were two indicators used to analyze the responses to CRRAT; the former referred to the situation wherein an individual looked at another's response and wrote down a similar answer, whereas the latter referred to when an individual gave a different answer in the same situation. This study calculated the counts of FO and IM in the two-player mode.

3. Results

3.1. Divergent thinking test

The two-way ANOVA results showed that the interaction effect between groups and one- and paired-player modes reached the level of significance in terms of fluency $[F(1, 340) = 26.28, p < .001, \eta^2 = 0.07]$, flexibility $[F(1, 340) = 11.06, p = .001, \eta^2 = 0.03]$, and originality $[F(1, 340) = 23.77, p < .001, \eta^2 = 0.07]$. The high- and low-score groups had significant main effects in terms of the scores for different indicators (Fs = 55.01, 55.79, 45.91, $ps < .001, \eta^2_s = 0.14, 0.14, 0.12$). Nonetheless, the scores for the two different modes showed no significant difference ($Fs = 1.32, 0.28, 0.01, ps = .251, 599, 981, \eta^2_s < 0.01$). Further analysis of the main effects indicated that the high-score group had a significantly better performance than the low-score group in terms of fluency [$F(1, 680) = 39.56, p < .001, \eta^2 = 0.00$], flexibility [$F(1, 680) = 31.88, p = .001, \eta^2 = 0.09$], and originality [$F(1, 680) = 33.34, p < .001, \eta^2 = 0.09$] in the single-player mode. Moreover, the high-score group had significantly higher scores than the low-score group [$Fs = 10.22, 8.14, 8.98, ps = .001, .002, .002, \eta^2_s = 0.03, 0.02, 0.03$] in the paired-player mode, but the effect size significantly decreased.

As shown in Table 1, the gap in AUT scores between the high- and low-score groups narrowed, as the low-score group improved the score [Fs = 7.90, 3.92, 11.77, ps = .003, .029, .001, $\eta^2_s = 0.02$, 0.01, 0.03] whereas the performance of the high-score group did not improve and even slightly declined [Fs = 19.70, 7.42, 12.00, ps < .005, $\eta^2_s = 0.05$, 0.02, 0.03]. In addition, the average flexibility and originality scores for each response were calculated with fluency as the denominator, and the corresponding results revealed that the average flexibility for the high-score group in the paired-player mode significantly improved [t(170) = 2.39, p = .018, Cohen's d = 0.18], whereas the average originality showed no noticeable changes [t(170) = -1.18, p = .242, Cohen's d = 0.09]. Conversely, the low-score group did not show any considerable changes in average flexibility [t(170) = 0.08, p = .940, Cohen's d = 0.01] and average originality [t(170) = 1.48, p = .141, Cohen's d = 0.11] in the paired-player mode. The results suggest that the interactive (i.e., paired-player) mode exerts a different impact on the high- and low-score groups in terms of divergent thinking.

Table 1. Descriptive statistics of two creativity tests between the two groups

_	High-score group		Low-score group	
	Mean	Standard deviation	Mean	Standard deviation
Divergent Thinking Test				
Fluency (single-player)	18.04	6.31	12.00	5.22
Fluency (paired-player)	16.22	7.02	13.15	6.43
Flexibility (single-player)	8.99	2.49	6.83	2.29
Flexibility (paired-player)	8.37	2.85	7.28	2.36
Originality (single-player)	17.02	8.95	9.67	6.46
Originality (paired-player)	15.25	9.51	11.43	8.09
Chinese Radical Remote Associates Test				
CCRAT (single-player)	.51	.15	.27	.15
CCRAT (paired-player)	.51	.18	.39	.2

Note. S = single-player; T = two-player; CCRAT = Chinese Radical Remote Associates Test. *N* for Divergent Thinking Test = 342; *N* for CCRAT = 318.

3.2. The CRRAT

The two-way ANOVA results showed that the groups and modes had significant interaction effects on the accuracy rate of the CCRAT [F(1, 316) = 35.82, p < .001, $\eta^2 = .10$]. Both the group [F(1, 316) = 111.91, p < .001, $\eta^2 = .26$] and the mode [F(1, 316) = 37.76, p < .001, $\eta^2 = .11$] had significant main effects. The analysis of the main effects indicated that the high-score group had a better performance than the low-score group, both in the single- [F(1, 632) = 73.81, p < .001, $\eta^2 = .19$] and two-player [F(1, 632) = 17.91, p < .001, $\eta^2 = .05$] modes, but the effect sizes declined. In addition, Figure 2 shows that the low-score group had a better performance in the two-player mode than in the single-player mode [F(1, 316) = 73.57, p < .001, $\eta^2 = .19$], whereas the high-score group showed no significant difference in both modes [F(1, 316) < .001, p = .99, $\eta^2 < .01$], suggesting that the gap between the low- and high-score groups narrowed due to the increase in score of the low-score group. This finding reveals that the interactive mode has no significant impact on the CRRAT performance of the high-score group.







3.3. Potential factors during interactive creativity

This study analyzed the strategies that individuals used when performing the AUT and CRRAT in the two-player mode. First, the low-score group ($M_{CC} = .45$, $SD_{CC} = .33$; $M_{PO} = .28$, $SD_{PO} = .22$) employed the CC and PO (ts = 2.39, 2.25, ps = .02, .03, Cohen's ds = .26, .24) more frequently than the high-score group ($M_{CC} = .37$, $SD_{CC} = .30$; $M_{PO} = .23$, $SD_{PO} = .21$) in the AUT. Relevant analysis results showed that the CC had a significant correlation with fluency (r = .12, p = .03) and originality (r = .12, p = .03) in the two-player mode, whereas it had no significant correlation with flexibility (r = .01, p = .80) in the two-player mode. However, the PO had no significant correlation with the above indicators in the two-player mode (rs = .04, -.06, .03, ps = .41, .27, .54).

Moreover, the low-score group ($M_{FO} = 4.13$, $SD_{FO} = 3.12$) had a greater number of FO counts [t(316) = 3.36, p = .001, Cohen's d = .38] in comparison with the high-score group ($M_{FO} = 3.04$, $SD_{FO} = 2.63$). However, the two groups had no noticeable differences in IM counts (t(316) = 0.87, p = .39, Cohen's d = .10). Relevant analysis results showed that the FO counts had a significant positive correlation with the two-player mode (r = .46, p < .001) and no significant correlation with IM counts (r = -.03, p = .55), suggesting that an individual is more likely to have a higher accuracy rate if they are more able to refer to another's response in the two-player mode.

4. Discussion

This study explores the differences in individuals' creative processes between divergent and convergent thinking in interactive situations. It employs an online interactive creativity task platform to gather participants' AUT and CRRAT performances in the single- and two-player modes, which differs from the previous studies that use penand-pencil tests to gather data about individuals' creativity performance. The online platform not only provides participants' scores for each dimension, but also records their question-answering process and responses. The results showed that those who obtained lower scores in the divergent thinking test in the single-player mode had a better performance in the two-player mode, especially on fluency, flexibility, and originality. Conversely, those who had a better AUT performance in the single-player mode had lower scores in the two-player mode, suggesting that the gap in the divergent thinking performance between the two groups had narrowed in interactive situations.

Moreover, the gap between the two members of a group for the CRRAT accuracy rate had also narrowed; individuals with a lower accuracy rate in the single-player mode improved their performance in the paired-player

mode, while others with a higher accuracy rate had no considerable changes in the paired-player mode. Additionally, this study found that in the paired-player mode, the more often an individual referred to others' responses, the more likely they would achieve higher scores on the divergent thinking test (especially for fluency and flexibility) and the CRRAT. These results reveal the similarity and difference between individuals' performance on the two creativity tests in interactive situations, and the possible correlation between their response strategies and their performances in the paired-player mode.

This study set the participants' performance in the single-player mode as their existing capability and found that the mode had interaction effects. The scores of those with high existing capabilities declined in fluency, flexibility, and originality on the divergent thinking test in the two-player mode. Conversely, those with low existing capabilities showed a considerable increase in the scores on the divergent thinking test in the two-player mode. The results reveal that performing the AUT in the two-player mode exerts different impacts on the high-and low-score groups. The performance of those with low existing capabilities may be improved due to their reference to the answers of those with high existing capabilities, whose answers give the low-score group inspiration and enable them to produce a greater number of more original ideas. However, the average flexibility score)/(Fluency score)) and average originality score ((Originality score)/(Fluency score)) of those with high existing capabilities in the two-player mode were not lower than their corresponding scores in the single-player mode. This decrease in their scores on the divergent thinking test may be attributed to their lower willingness to answer questions when the other group members referred to their ideas, which affected their fluency.

What is worth mentioning is that the extent to which both the low-score group improved their performance and the scores of the high-score group declined was different. In other words, in the paired-player mode, those with high capabilities still had a better performance than those with lower scores, even though they improved the low scorers' divergent thinking by referring to the ideas of those with high capabilities, suggesting that the collaboration in the paired-player mode only partially improved the divergent thinking performance of those with low capabilities.

In addition, individuals' CCRAT scores in the two-player mode significantly increased. Further examination on the impact of the mode on participants with different existing capabilities found that in the two-player mode, only those with low existing capabilities improved their performance, whereas those with high existing capabilities had no change in their scores. This finding reveals two phenomena. First, those with low existing capabilities can refer to another's response when performing the CRRAT in the two-player mode. This enables them to come up with an answer that they are not able to think of independently and to spend less time answering certain test questions when they are able to refer to another's, which allows them to have more time for other questions and to come up with better answers, thus improving their performance in the two-player mode, indicating that those with low existing capabilities brought limited assistance to them. Moreover, individuals could refer to another's response when performing the CRRAT consisting of close-ended questions, but the responses of the high- and low-score groups were not completely the same. This finding suggests that individuals do not completely refer to another's response but may refer to another's response strategy, resulting in differences between the two groups.

In sum, this study found that the paired-player mode exerted different impacts on one's performance on the divergent thinking test and CRRAT, especially for those with high capabilities. The AUT and CRRAT can be differentiated in terms of task types (Wakefield, 1992). The AUT is composed of a close-ended question with open-ended solutions, whereas the CRRAT consists of open-ended questions of fixed patterns that have close-ended answers. The AUT respondents freely associated the task with target-related concepts, which can be explained via the associative hierarchy (Mednick, 1962). Highly creative people have a greater chance to produce more original ideas, while those with low creativity tend to come up with fewer creative ideas. However, the AUT sets time limits on one's response, but no limits on the number of their solutions. In interactive situations, those with high capabilities may have lowered their willingness to think of more solutions after considering that their ideas may be referred to, which resulted in their fluency score decreasing in the paired-player mode than in the single-player mode.

However, their average flexibility and originality scores in the two modes remained unchanged, suggesting that their creativity performance did not get worse in the paired-player mode. Those with high capabilities in the CRRAT did not slightly decrease as well. They were required to finish the same number of test questions in the single- and paired-player modes, and put on the same performance. This may result from the ceiling effect of one's creative performance, or the shortage of motivation to deliver a better performance or reference target. In

sum, the paired-player mode facilitates the improvement of the creativity performance for those with low capabilities, but has limited effects on those with high capabilities.

This study further explores the relationship between potential factors and one's performance in the paired-player mode. In terms of the divergent thinking test, a high percentage of those with low capabilities referred to the response of the other members in the paired-player mode, thus producing more unique ideas. In addition, the CC rate was positively correlated with the fluency and originality scores in the paired-player mode, while the PO rate had no correlation with divergent thinking. In interactive situations, the way one referred to the other (i.e., FO or IM) was conducive to divergent thinking, but the original ideas produced this way were independent from the scores of the divergent thinking test.

In this respect, those with low capabilities often referred to the answer of the other group members in the CRRAT (i.e., FO), which had a positive correlation with the accuracy rate, whereas the IM had no correlation with the accuracy rate. This finding indicates that FO is more helpful to lift the accuracy rate of one's CRRAT (close-ended questions). These results also suggest that, generally, FO exerts a positive impact on one's creativity performance in the paired-player mode, which means that one is able to produce more original ideas and is more likely to solve remote associates' problems if referring to others' responses in the paired-player mode and gaining inspiration. However, this synergetic effect only occurred for those with low capabilities in this study.

4.1. Limitations and future research

This study has some limitations regarding its implementation. First, the online interactive creativity task platform only enabled participants to access the response of their group members before coming up with other solutions in the paired-player mode. This function sufficed when conducting this study, However, technically speaking, it did not enable the two group members to interact with each other during the question-answering process, which means that two group members could not talk to each other and discuss test questions (Pifarré, 2019). Thus, this platform remains to be improved in subsequent research.

In addition, this study found that the AUT performance of the high-score group in the single-player mode slightly declined in the paired-player mode, potentially because they were not happy to see the fruit of their labor being plagiarized, which reduced their motivation to deliver a good performance on the divergent thinking test. This speculation needs to be verified by more empirical experience via interviews or the manipulation of modes (like cooperative or competitive modes). The high-score group in the single-player mode did not deliver a better performance in the paired-player mode, possibly because of a limited response time, participants' motivation to answer questions, and the ceiling effect of one's creativity. The possible causes cannot be verified in this study, so they remain to be confirmed in future studies.

5. Conclusion

This study is the one of the first to use an online interactive creativity task platform to discuss the difference in the creativity performance between two modes. The comparison of the high- and low-score groups' AUT and CRRAT performances in the two-player mode examined the impact of the two-player mode on creativity. The results showed that the low-score group refer to the response of the other group members and significantly improved their creativity performance in the two-player mode, while the high-score group did not significantly increase their AUT and CRRAT scores. Further analysis of the impact of the two-player mode on the divergent and convergent creativity revealed that the CCRAT accuracy rate increased while the AUT scores did not show a noticeable increase. In short, one plus one is not greater than two under any circumstances, which may be attributed to the mode.

The results reflect the importance of mutual observation between group members in the process of group creation. Teachers can guide students to open their minds and learn from group members, so that students with different capacities can come up with a variety of ideas through observing and imitating ideas of other group members. Furthermore, this study analyzed how people with different capacities might improve their creative performance by referring to the others' answers. This study adopted an online standardized measurement tool to conduct preliminary research on how two participants perform on creativity tests in a concerted effort, which expands the potential contributions of co-creativity research.
Acknowledgement

This work was financially supported by the "Institute for Research Excellence in Learning Sciences" of National Taiwan Normal University (NTNU) from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan. We also thank the Ministry of Science and Technology, Taiwan, R.O.C. for funding this study, through projects on "How is one plus one more than two? Exploring the process and training of creative thinking and creative product under interactive situations: From measurement, mechanisms to neural plasticity" (MOST108-2410-H-003-080) and "How do individuals display their creativity in groups? The moderation effects of intrinsic and extrinsic factors, brain structure and interactive mobile teaching" (MOST109-2628-H-003-005-MY2).

References

Amabile, T. M. (1988). A Model of creativity and innovation in organization. *Research in Organizational Behavior*, 10, 123-167.

Bowden, E. M., & Jung-Beeman, M. (2003). Normative data for 144 compound remote associates problems. *Behavior Research Methods, Instruments, and Computers, 35*, 634-639. https://doi.org/10.3758/BF03195543

Bronfenbrenner, U. (1979). The Ecology of human development: Experiments by nature and design. Harvard University Press.

Chang, J. H., Chen, H. C., Hsu, C. C., & Lin, Y. N. (2009). 高中職創意競賽之團隊歷程一成員歧異度與團隊氣氛對團隊 創造力的影響 [A Study of team creativity process of high school students in Taiwan – The Impact of the team climate and team diversity on team creativity]. *Journal of Education & Psychology*, *32*(4), 73-97.

Chang, Y. L., Wu, J. Y., Chen, H. C., & Wu, C. L. (2016). 「中文部件組字遠距聯想測驗」之發展與其信效度研究 [The Development of Chinese radical remote associates test]. *Psychological Testing*, 63(1), 59-81.

Clapham, M. M. (2010). The Effects of affect manipulation and information exposure on divergent thinking. *Creativity Research Journal*, 13, 335-350. https://doi.org/10.1207/S15326934CRJ1334_11

Csikszentmihalyi, M. (1988). The Flow experience and its significance for human psychology. In M. Csikszentmihalyi & I. S. Csikszentmihalyi (Eds.), *Optimal Experience: Psychological Studies of Flow in Consciousness* (pp. 15–35). Cambridge University Press.

Fleck, J. S., & Weisberg, R. W. (2013). Insight versus analysis: Evidence for diverse methods in problem solving. *Journal of Cognitive Psychology*, 25, 436-463. https://doi.org/10.1080/20445911.2013.779248

Glăveanu, V. P., Gillespie, A., & Karwowski, M. (2018). Are people working together inclined towards practicality? A Process analysis of creative ideation in individuals and dyads. *Psychology of Aesthetics, Creativity, and the Arts, 13*(4), 388-401. https://doi.org/10.1037/aca0000171

Guilford, J. P. (1956). The Structure of intelligence. Psychological Bulletin, 53, 267-293. https://doi.org/10.1037/h0040755

Hong, J. C., Wu, C. L., Chen, H. C., Chang, Y. L., & Chang, K. E. (2016). Effect of radical-position regularity for Chinese orthographic skills of Chinese-as-a-second-language learner. *Computers in Human Behavior*, 59, 402-410. https://doi.org/10.1016/j.chb.2016.02.002

Hsu, C. C., Chen, H. C., & Chiu, F. C. (2012). 「報紙的不尋常用途」測驗之編製 [The Development of unusual uses of the newspapers test]. *Journal of Chinese Creativity*, 3(2), 33-56.

Huang, P. S. (2017). An Exploratory study on remote associates problem solving: Evidence of eye movement indicators. *Thinking Skills and Creativity*, 24, 63-72. https://doi.org/10.1016/j.tsc.2017.02.004

Huang, P. S., Chen, H. C., & Liu, C. H. (2012). 「中文詞彙遠距聯想測驗」之編製及其信、效度報告 [The Development of Chinese word remote associates test for college students]. *Psychological Testing*, 59(4), 581-607.

Huang, P. S., Liu, C. H., & Chen, H. C. (2019). Examining the applicability of representational change theory for remote associates problem solving with eye movement evidence. *Thinking Skills and Creativity*, *31*, 198-208. https://doi.org/10.1016/j.tsc.2018.12.001

Hung, S. P., & Wu, C. L. (2021). Cognitive component analysis comparing three Chinese remote associates Tests: Linear logistic latent trait model approach. *Creativity Research Journal*, 33(3), 224-234. https://doi.org/10.1080/10400419.2021.1872287

Jen, C. H., Chen, H. C., Lien, H. C., & Cho, S. L. (2004). 創造力測量的輔助工具:中文遠距聯想量表的編製 [The development of the Chinese remote association test]. *Research in Applied Psychology*, 21, 195-217.

Kaufmann, G. (2003). Expanding the mood-creativity equation. *Creativity Research Journal*, 15, 131-135. https://doi.org/10.1207/S15326934CRJ152&3_03

Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An Eye movement study of insight problem solving. *Memory & Cognition*, 29, 1000-1009. https://doi.org/10.3758/BF03195762

Leonard, D. A., & Swap, W.C. (1999). When sparks fly: Igniting creativity in group. Harvard Business School Press.

Lin, W.-L., & Lien, Y.-W. (2013). The Different roles of working memory in open-ended versus closed-ended creative problem solving: A Dual-process theory account. *Creativity Research Journal*, 25, 85-96. https://doi.org/10.1080/10400419.2013.752249

Lin, W.-L., Hsu, K.-Y., Chen, H.-C., & Wang, J.-W. (2012). The Relations of gender and personality traits on different creativities: a dual-process theory account. *Psychology of Aesthetics, Creativity and the Arts, 6*, 112-123. https://doi.org/10.1037/a0026241

Lin, W.-L., Lien, Y.-W., & Jen, C.-H. (2005). 想得多是想得好的前提嗎?探討發散性思考能力在創意問題解決的角色 [Is the More the Better? The Role of divergent thinking in creative problem solving]. *Chinese Journal of Psychology*, 47, 211–227.

Littleton, K., Rojas-Drummond, S., & Miell, D. (2008). Introduction to the special issue: 'collaborative creativity': Sociocultural perspectives. *Thinking Skills and Creativity*, 3(3), 175–176. https://doi.org/10.1016/j.tsc.2008.09.004

Magyari-Beck, I. (1993). Creatology: A Postpsychological study. *Creativity Research Journal*, 7, 183-192. https://doi.org/10.1080/10400419409534523

Mednick, S. A. (1962). The Associative basis of the creative process. *Psychological Review*, 44, 220-232. https://doi.org/10.1037/h0048850

Mednick, S. A. (1968). The Remote associates test. Journal of Creative Behavior, 2, 213-214.

Pifarré, M. (2019). Using interactive technologies to promote a dialogic space for creating collaboratively: A Study in secondary education. *Thinking Skills and Creativity*, 32, 1-16. https://doi.org/10.1016/j.tsc.2019.01.004

Rhodes, M. (1961). An Analysis of creativity. Phi Delta Kappa, 42, 305-310.

Sternberg, R. J., & Lubart, T. I. (1999). The Concept of creativity: Prospect and paradigms. In R. J. Sternberg (Ed.), Handbook of Creativity (pp. 3-15). Cambridge Press.

Torrance, E. P. (1974). The Torrance tests of creative thinking: Norms-technical manual. Personal Press.

Wakefield, J. F. (1992). Creative thinking: Problem solving skills and the art orientation. Ablex.

Wallas, G. (1926). The Art of thought. Harcourt, Brace.

Walsh, C., Chappell, K., & Craft, A. (2017). A Co-creativity theoretical framework to foster and evaluate the presence of wise humanising creativity in virtual learning environments (Vles). *Thinking Skills and Creativity*, 24, 228-241. https://doi.org/10.1016/j.tsc.2017.01.001

Weisberg, R. W. (2015). Toward an integrated theory of insight in problem solving. *Thinking & Reasoning*, 21(1), 5-39. https://doi.org/10.1080/13546783.2014.886625

Woodman, R. W., Sawyer, J. E., & Griffin, R. W. (1993). Toward a theory of organizational creativity. *The Academy of Management Review*, 18(2), 293-321. https://doi.org/10.2307/258761

Wu, C. L. (2019). Discriminating the measurement attributes of the three versions of Chinese remote associates test. *Thinking Skills and Creativity*, *33*, 100586. https://doi.org/10.1016/j.tsc.2019.100586

Wu, C. L., Huang, S. Y., Chen, P. Z., & Chen, H. C. (2020a). A Systematic review of creativity-related studies applying the remote associates test from 2000 to 2019. *Frontiers in Psychology*, *11*, 573432. https://doi.org/10.3389/fpsyg.2020.573432

Wu, C. L., Peng, S. L., & Chen, H. C. (2021). Why can people effectively access remote associations? Eye movements during Chinese remote associates problem solving. *Creativity Research Journal*, 33(2), 158-167. https://doi.org/10.1080/10400419.2020.1856579

Wu, C. L., Tsai, M. N., & Chen, H. C. (2020b). The neural mechanism of pure and pseudo-insight problem solving. *Thinking & Reasoning*, 26(4), 479-501. https://doi.org/10.1080/13546783.2019.1663763

Zeilig, H., West, J., & van der Byl Williams, M. (2018) Co-creativity: Possibilities for using the arts with people with a dementia. *Quality in Ageing and Older Adults, 19*(2), 135-145. https://doi.org/10.1108/QAOA-02-2018-0008

Guest Editorial: Blockchain in Smart Education

Ching-Hsien Hsu^{1,2,3*}, Amir H. Alavi⁴ and He Li⁵

¹Department Computer Science and Information Engineering, Asia University, Taiwan // ²Department of Medical Research, China Medical University Hospital, China Medical University, Taiwan // ³Guangdong-Hong Kong-Macao Joint Laboratory for Intelligent Micro-Nano Optoelectronic Technology, School of Mathematics and Big Data, Foshan University, Foshan, China // ⁴Department of Civil and Environmental Engineering, Department of Bioengineering, Swanson School of Engineering, University of Pittsburgh, USA // ⁵Department of Sciences and Informatics, Muroran Institute of Technology, Japan // chh@cs.ccu.edu.tw // ALAVI@pitt.edu // heli@mmm.muroran-it.ac.jp

*Corresponding author

ABSTRACT: In a smart educational environment, the significant challenges faced by its stakeholders are trust, privacy, and transparency-related issues in sharing and retrieval of any information. Since blockchain is a sole technology provides extraordinary features such as decentralization, traceability, and immutability; integrating this technology in a smart educational environment it can overcome all the technical risks, potential threats, and privacy concerns. This special issue aims at presenting the current state-of-the-art research and future trends on various aspects of the implementation of blockchain technologies that address the gaps prevailing in copyright and credential related issues, which can be seen as a promising sign for any virtual or innovative developments in teaching and learning platform. Papers selected for this special issue represent recent progress in the field, including works on VR Assisted Learning Environment, Secure Scoring Mechanism for Online Learning, Psychological Distance of Online Education.

Keywords: Blockchain, Smart education, Virtual reality, Online learning, Education theory

1. Introduction

Blockchain is one of the ingenious technologies which are disrupting the future of many industries. This encrypted digital ledger technology has all the potential to reshape areas such as healthcare, education, and finance. Education is one such area where these blockchain-based techniques and properties can trigger a wide range of opportunities. In a smart educational environment, the significant challenges faced by its stakeholders are trust, privacy, and transparency-related issues in sharing and retrieval of any information. Since blockchain is a sole technology provides extraordinary features such as decentralization, traceability, and immutability; integrating this technology in a smart educational environment it can overcome all the technical risks, potential threats, and privacy concerns. Whether the educational environment is formal or informal the data can be stored and accessed more securely by using blockchain appropriately. Moreover, the application of blockchain in a smart educational system shall also provide smart assistance for implementation, evaluation, tracking, delivery, and management of any information concerning both the teacher and the learner.

Due to the huge volumes of educational data across various learning platforms, the protection of sensitive and valuable information needs the embracement of robust and intelligent technology. This leads to the development of a decentralized distributed blockchain technology, where each node is secured by a blockchain ledger which can be accessed only by the private key. Furthermore, the principal advantage of the blockchain technology is that the information is stored within the blockchain network with a unique identity, so that when the information is accessed by the users it is checked and validated properly by comparing all the related data. On the other hand, Smart Contracts is a traceable digital transaction facilitator used along with the blockchain technology in a smart educational environment could make the overall system more secure, reliable and more transparent.

This special issue aims at presenting the current state-of-the-art research and future trends on various aspects of the implementation of blockchain technologies that address the gaps prevailing in copyright and credential related issues, which can be seen as a promising sign for any virtual or innovative developments in teaching and learning platform. The main areas covered by this special issue or main topics cover methodologies, modeling, analysis and newly introduced applications. Besides the latest research achievements, this special issue also deals with innovative commercial management systems, innovative commercial applications of educational technology, and experience in applying recent research advances to real-world problem.

Papers selected for this special issue represent recent progress in the field, including works on VR Assisted Learning Environment, Secure Scoring Mechanism for Online Learning, Psychological Distance of Online Education. All of these papers not only provide novel ideas and state-of-the-art techniques in the field, but also stimulate future research in the sustainable learning environments.

2. VR assisted learning

In order to conform to the glacier terrain environment and textbook knowledge, the system was continuously discussed and revised with the high school geography teachers during the development process to ensure that the terrains observed by students conform to the textbook teaching materials and the actual teaching content. The paper by Chen and Chen (2022), entitled "Exploring the Effect of Spatial Ability and Learning Achievement on Learning Effect in VR Assisted Learning Environment," constructed a teaching software of glacier terrain, which allows students to explore freely in virtual environment and the effect of different learning modes. This study also collected students' spatial ability and geographical learning achievement, and explored whether students' spatial ability and geographical learning achievement affects their learning effect. Through statistical analysis, it is found that "Spatial Visualization" in students' spatial ability positively affected their learning performance in virtual reality software. This study verifies the value of using virtual reality to assist geography course learning, and provides more references and suggestions for future researchers.

3. Secure online learning

With the rapid increase in online learning and online degree programs, the need of secure and fair scoring mechanisms for online learning becomes urgent. The paper by Tsai et al. (2022), entitled "Design and Development of a Secure Scoring Mechanism for Online Learning Based on Blockchain," designed and developed a secure scoring mechanism based on blockchain to build transparent and fair interactions among students and teachers. The proposed scoring mechanism was implemented by employing Ethereum and its three autonomous smart contracts, and the robustness was also verified by experiments to prove the feasibility of the system. this system helps manage interactions among students and teachers during the process of educational assessment, and encourages all on-chain members to trust the online learning process. These system features also help conduct peer evaluation and self-management that are essential for a student-centered and collaborative learning environment, for which is what emerging educational trend advocates.

4. Online education theory

In education reform, people actively promote education innovation through the application of intelligent technology. Especially, blockchain in smart education has the technical characteristics of peer-to-peer transmission, data fidelity, intelligent contract, etc., which provides a feasible technical scheme for educational reform. The paper by Zhang et al. (2022), entitled "A Grounded Theory Research on the Psychological Distance of Online Education," reconstructed the theoretical model of psychological distance in the process of online education. The authors investigated the psychological perception of online education users and combine psychological distance and grounding theory. Findings are (1) The key to online education; (3) Factors influencing psychological distance in online education; (4) Resulting evaluation of online education – psychological perception evaluation.

5. Conclusions

All of the above papers address either technical issues in educational technologies or information security or propose novel application models in the various smart-learning systems and social computing fields. They also trigger further related research and technology improvements in application of novel educational technologies. Honorably, this special issue serves as a land-mark source for education, information, and reference to professors, researchers, and graduate students interested in updating their knowledge of block chain, cyber-physical-system, augmented reality, and novel application models for future on-line learning and teaching systems.

The special issue of this journal covers different aspects of the problem, from both the theoretical and the practical side. After a large open call, an international editorial committee selected three research papers.

Acknowledgement

The guest editors would like to express sincere gratitude to Prof. Maiga Chang for giving the opportunity to prepare this Special Issue. In addition, we are deeply indebted to numerous reviewers for their professional effort, insight, and hard work put into commenting on the selected articles that reflect the essence of this special issue. We are grateful to all authors for their contributions and for undertaking two-cycle revisions of their manuscripts, without which this special issue could not have been produced.

References

Chen, C.-C., & Chen, L.-Y. (2022). Exploring the effect of spatial ability and learning achievement on learning effect in VR assisted learning environment. *Educational Technology & Society*, 25(3), 74-90.

Tsai, C.-T., Wu, J.-L., Lin, Y.-T., Yeh, M. K.-C. (2022). Design and development of a Blockchain-based secure scoring mechanism for online learning. *Educational Technology & Society*, 25(3), 105-121.

Zhang, L., Ye, J., & Wang, J. (2022). A Grounded theory study of the psychological distance in online education. *Educational Technology & Society*, 25(3), 91-104.

Exploring the Effect of Spatial Ability and Learning Achievement on Learning Effect in VR Assisted Learning Environment

Chia-Chen Chen* and Liang-Yu Chen

Department of Management Information Systems, National Chung Hsing University, Taiwan // emily@nchu.edu.tw // g106029003@mail.nchu.edu.tw

*Corresponding author

ABSTRACT: Based on "Virtual Reality" (VR), this study constructed a teaching software of glacier terrain, which allows students to explore freely in virtual environment and the effect of different learning modes. This study also collected students' spatial ability and geographical learning achievement, and explored whether students' spatial ability and geographical learning achievement affects their learning effect. In this study, senior high school sophomores were selected as the experimental subjects. The students involved were divided into the experimental group and the control group. The experimental group students used VR glacier terrain teaching software to assist their course learning, while the control group students learned the geography course by means of traditional teaching methods. After the experiment, it is found that students in the experimental group performed better than those in the control group in the post-test, and students with high spatial ability and high geographical achievement performed better in the post-test. Through statistical analysis, it is found that "Spatial Visualization" in students' spatial ability positively affected their learning performance in virtual reality software. In the interviews and feedback sheets after the experiment, most students had positive attitudes towards using virtual reality assisted software for geography course learning. This study verifies the value of using virtual reality to assist geography course learning, and provides more references and suggestions for future researchers.

Keywords: Virtual reality, Spatial ability, Geographical learning, Learning achievement, Learning effect

1. Introduction

Among many technologies, virtual reality technology, widely discussed in recent years, has great potential for digital learning. Virtual reality (VR) has been growing substantially both in the aspects of technology and market since 2016, which is also referred to as the first year of VR. According to the survey report, spending on virtual reality and augmented reality will grow from just over \$12.0 billion in 2020 to \$72.8 billion in 2024 (International Data Corporation, 2020), and will also be gradually popularized in the market. VR technology can provide users with a borderless virtual environment, allowing people to immerse themselves in it, whereby we can experience different scenes without going on a long journey (Woodford, 2019). Nowadays, many studies have incorporated virtual reality technology into the field of digital learning (Chen et al., 2020; Hsiao et al., 2020; Ip et al., 2018; Lai et al., 2021; Lin, 2017). This study intended to assist senior high school students in learning geography through the features of virtual reality, and to explore whether this learning method can improve students' learning effect and motivation, as well as to understand the problems which might be encountered in virtual reality-assisted course. Researchers who also conduct virtual reality-assisted learning experiments in the future will get useful experience and suggestions from this study.

This study cooperated with a senior high school in Taichung to jointly develop with the senior high school teachers a virtual reality-assisted learning software for the "New Zealand and Australia" unit of the geography course in senior high sophomore. The software was introduced into the geography course in the first semester of sophomore year, allowing students to feel their "Personal Experience" brought by virtual reality. Different from the traditional geography learning course in the past, the new learning mode can deepen students' impression on geographical knowledge. Past literature pointed out that learning through virtual reality can improve students' learning effect and reduce risk factors (Chen, & Huang, 2020; Limniou et al., 2008). Taking this study as an example, students can observe the magnificent glacier terrain types in the classroom, avoid rugged and dangerous terrains and lethal low temperature in the glacier terrains, and use the features of virtual reality technology to assist the teaching of courses, studies on the use of immersive virtual reality technology applied to senior high school courses are still rare. Therefore, this study hopes to improve the current teaching quality of senior high school geography education, and allow basic education to keep up with the evolution of science and technology.

Students' spatial ability is also a very important factor in the process of exploring virtual reality-assisted learning. Relevant studies have pointed out that students' spatial ability affects their learning effect in virtual

reality environment. The results show that students with different spatial ability have different learning performance (Lee & Wong, 2014). However, due to the lack of discussion on the relationship between "Spatial Ability" and "Immersive" virtual reality assisted learning, most of the studies focus on "Non-immersive" virtual reality technology (Ip et al., 2018; Lin, 2017; Vishwanath et al., 2017). Therefore, this study not only explores the learning effect of virtual reality assisted learning, but also expects to understand whether students with different spatial ability and learning achievement have different influences on immersive virtual reality learning.

"Learning Achievement" refers to students' relevant knowledge in a field they have acquired prior to the study is conducted, and their personal experience that can influence the experiment and the study results. Namely, before the course experiment is conducted, will students' learning achievement affect their learning performance and learning outcome? This study aimed at assisted learning of geography course in senior high school, expecting to understand whether students with different "Learning Achievement" have different learning effect. This study utilized the students' final grades of geography in the second semester of freshman year as students' "Learning Achievement" scores, and formulated a more appropriate teaching method through observing the effect brought by students' "Learning Achievement." This study expects to explore whether virtual reality assisted learning help students improve their learning effect. In the other word, the "Learning Effect" means the learning performance like the post-test score of learners.

In this era of information explosion, new learning models will inevitably become the future trend. This research hopes to explore whether Taiwanese students use virtual reality to assist learning, compared with traditional methods, to improve learning effectiveness and understand different spaces. Whether students with ability and learning achievement will have an impact on their learning performance, after the above research objectives, we sorted out the following research objectives as the main topics to be explored in this research.

- Construct a VR-assisted learning system to improve students' learning method and increase learning effectiveness.
- Understand the impact of spatial ability on learning performance
- Understand the impact of learning achievement on learning performance

2. Literature review

The system developed in this study is based on the three main themes of immersive virtual reality, spatial ability and geography course, of which the software of geography course was developed with the features of virtual reality. The system also verified the performance of students with different spatial ability and learning achievement, explored whether virtual reality-assisted learning course and traditional learning method affect students' learning effect and compared the differences therein.

2.1. Virtual reality

Virtual reality technology has been widely used in our lives in recent years. Traces of virtual reality can be found from virtual museums (Carvajal et al., 2020), simulation of experimental courses to game entertainment (Huang et al., 2016). "Virtual Reality" is a kind of virtual environment transformed from people's imaginary environment through computer operation and picture presentation by using computer technology. This virtual environment is called virtual reality because it is like real existence to people's sensory organs. At present, many studies have introduced "Virtual Reality" technology into the teaching courses (Lee, & Wong, 2014; Lovreglio et al., 2017) to explore whether the use of VR technology improves students' learning effect and motivation. In the following, we continue to discuss the relevant studies on integrating "Virtual Reality" technology into the teaching courses in recent years.

With the features of virtual reality mentioned above, virtual reality can build a highly realistic and low-risk teaching environment. Some studies based on VR features implemented "Fire Escape Training System" and "Earthquake Escape Training System," allowing users to carry out zero-risk escape training by using virtual reality devices. In virtual reality environment, trainees can be assured that they are not harmed by fires and earthquakes and can experience disasters in a highly realistic environment (Lovreglio et al., 2017). In the aspect of cost reduction, VR also has excellent performance. Some studies used virtual reality in combination with frog anatomy courses; although each anatomy course requires many living frogs (Lee & Wong, 2014), students can repeat the operation of anatomy courses without using living frogs in virtual reality environment, which greatly

reduces the cost of frog anatomy courses, and provides students with the opportunity to repeat operations, thereby improving students' skillfulness and proficiency.

The literature pointed out that in order to use virtual reality to teach students, it is suggested that students should operate from the first person perspective and in a joystick way (Lovreglio et al., 2017). Compared with the fixed perspective presentation, this mode of operation is helpful to improve students' spatial memory, whereby increasing students' learning effect. This mode is also the very operation method adopted by this study. Bashabsheh et al. (2019) used VR software to construct the 4D model (3D model and time dimension) for certain building construction phases to do immersive and non-immersive virtual reality experience for the learners. Ip et al. (2018) explored whether virtual reality learning can help students with Asperger's disease improve their interpersonal communication ability; the results show that giving social training to students with Asperger's disease through immersive virtual reality training can make students understand other people's emotions more easily and improve their ability to interact with others. Lin (2017) explored the coping strategies of men and women when facing the VR horror games. The results showed that the strategies of men and women are very different. Men mostly deal with them in the way of facing and fighting back, while women mostly avoid and escape. Vishwanath et al. (2017) adopted VR to teach history and geography for low-income children in India. The results showed that virtual reality can significantly improve students' learning motivation. Huang et al. (2016) investigated students' attitudes towards using virtual reality in pharmacy course. The results showed that students have a positive attitude towards using virtual reality technology to assist pharmacy course learning.

As most of the digital learning fields are currently assisted by non-immersive virtual reality and semi-immersive software. Few literature discuss whether immersive virtual reality technology improve students' learning effect. This study used immersive virtual reality technology to explore whether the learning effect of senior high school students in geography course is significantly improved. The greatest influence of immersive and non-immersive virtual reality on this study is to be able to use the first person perspective to observe the glacier terrains, allowing students to have sufficient senses of being introduced, attract students' attention and enhance their learning motivation by using new technology.

2.2. Spatial ability

The current academic definition of spatial ability has not been a general consensus agreed by scholars. Some scholars believe that there are three main parts of spatial ability, namely, "Spatial Visualization," "Mental Rotation" and "Spatial Perception" (Linn & Petersen, 1985), wherein spatial visualization refers to the ability to transform complex spatial information into required information, mental rotation refers to the ability of people to rotate and move objects in their brains so as to observe different orientations of objects, while spatial perception refers to the ability to understand spatial information and distinguish between correct and incorrect spatial information.

The spatial ability test book used by this study was the "Academic Aptitude Test Book" which developed by College Entrance Examination Center for senior high school students designed by the College Entrance Examination Center of the Republic of China. This test book divides spatial ability into three main facets: "Spatial Perception," "Spatial Visualization" and "Spatial Positioning," and tests students' individual spatial ability through three different types of questions.

From the past literature (Donnon et al., 2005; Merchant et al., 2013; Metoyer et al., 2015), we can know that Spatial Perception is defined as the ability to collect and analyze information around oneself, and to perceive spatial information about one's own body position, Spatial Visualization is defined as the ability to manipulate and rotate two-dimensional images and three-dimensional graphics in the mind through imagination and spirit, to perceive images and graphics from different angles through imagination, while Spatial Orientation is defined as the ability of the human brain to allow people to control their own heading toward destinations and maintain body balance and function in the course.

From the above-mentioned literature, we can know that students' spatial ability affects significantly many subjects' learning, and in the environment of virtual reality assisted instruction, spatial ability is also an essential factor frequently discussed. This experiment incorporated immersive virtual reality technology with geography course. Although many studies have been carried out on the correlation between virtual reality and spatial ability, studies exploring the correlation between "immersive" virtual reality and spatial ability are still quite rare. Therefore, whether students' spatial ability affects their learning effect by using immersive virtual reality assisted devices is the focus of this study.

Recently, Tchoubar et al. (2018) survey whether spatial ability affects students' learning on digital learning platform. The results showed that spatial ability affects students' ability to use digital platform significantly. Roch et al. (2018) studied the influence of spatial ability on laparoscopic navigation training. The results show that spatial ability positively affects the results of laparoscopic navigation training. Tascón et al. (2018) explored whether testees of different ages affect their performance in virtual reality. The results presented that the older the testees are, the worse their performances are. In virtual reality, the testees would perform better if they operate with walking mode.

As the above literature, previous researches discuss the spatial ability will influence the students' learning effect on e-learning platform and laparoscopic navigation training. Therefore, this study designed the geography course with 3D VR technique. The goal aims to investigate the different degree of spatial ability of students would have the different learning achievements by using new technology.

2.3. Geographical subjects

Many scholars have been devoted themselves to introducing digital technology into the basic education courses, hoping to improve students' learning efficiency and effect. "Geography Course" is often regarded as one of the main items of digital learning. Some literature pointed out that elementary school students can significantly improve their learning by using "Game-based" virtual reality software to assist students in learning geography course, which elevates significantly children's learning motivation and learning effect, and train children to learn independently (Chen et al., 2017; Lemke et al., 2000).

In another literature, virtual reality technology was used to assist elementary school students in learning geography course (Chen et al., 2017; Lemke et al., 2000); in this study, it is found that students' learning motivation and learning effect can be improved by using virtual reality to learn geography course against the traditional teaching methods. We can conclude from the above-mentioned literature that "Virtual Reality" plays an indispensable role in the field of digital learning nowadays.

It has been pointed out in a literature that there is no difference between the traditional learning method and learning with tablet devices in the learning effect of geography course. However, if the two teaching modes are combined in the process of geography course, the individual shortcomings of each teaching method can be made up, thus improving students' learning effect (Walczak & Taylor, 2018). In traditional geography teaching, the teaching materials used by students are often flatly printed, and the maps used are often ill-designed. In fact, it is difficult for students to construct perfect geography concepts by using books alone. Therefore, many scholars believe that introducing virtual reality technology into geography course can help students better understand the geographical information in books (Lin et al., 2013), while in the immersive virtual reality teaching, some scholars used Google Cardboard as the learning medium to assist elementary school students in learning the seven wonders of the world; although students' learning effect was not been verified in the experiment, it was felt clearly that students have positive attitudes towards the introduction of new technology into the course (Lv, & Li, 2015). Some studies have designed a virtual reality assisted learning course for elementary school geography course, incorporating visits and surveys of reality geographical environment and collecting data through taking pictures. The collected data were incorporated with course knowledge to construct a virtual environment that conforms to the actual terrains and deepens students' impression on the teaching content. The research results indicated that after the virtual reality assisted learning course, students' geographical level significantly improved, and more courses designed with different themes were expected (Hsu, & Chan, 2018).

Looking at many of the above-mentioned studies on introducing digital learning into geography courses, many scholars have introduced virtual reality technology into geography courses. Because the connotation of geography is actually closely related to the environment around us, virtual reality has various presentations. The characteristics of various environments, the combination of geography courses and virtual reality is a complementary application, but most of the current research focuses on the research of "non-immersive" virtual reality technology, and few scholars explore the use of "immersive" virtual reality technology. Therefore, this research was the focus of whether VR technology will affect students' learning efficiency and their spatial ability on geography courses.

3. Software development and system implementation

In order to conform to the glacier terrain environment and textbook knowledge, the system was continuously discussed and revised with the high school geography teachers during the development process to ensure that the terrains observed by students conform to the textbook teaching materials and the actual teaching content. Due to the individuality of virtual reality in the process of one's using software, the system must have the function of automatically recording students' exploratory behaviors so that after students complete the software operation, teachers and developers can understand students' performance in virtual reality environment through the recorded data of the system. This study used immersive virtual reality technology different from that in most previous literature to develop teaching assisted software, and expected to improve students' learning effect through this software; it also explored the problems and situations faced by students when using immersive virtual reality technology to assist in learning the course.

3.1. System design

The system used in this study is a self-developed "Virtual Reality Assisted Learning Device (VALID)," which assists students in learning glacier terrains. The system needs the connection and communication among the three devices of "VR Box" virtual reality glasses, smart mobile phones and "RoHS" Bluetooth remote control gamepad. However, the VR also bring negative effects, such as sickness or virtual fatigue. Therefore, when using VALID, students first have to start the "Virtual Reality Assisted Instruction Software" in the smart mobile phone and put the smart mobile phone into the VR Box, and then use the VR Box to adjust the focal length and the tightness of the head wearing so that students can comfortably use VALID. Next students have to turn on the pre-set RoHS Bluetooth gamepad and let the Bluetooth gamepad connect with the smart mobile phone. After the Bluetooth gamepad is connected with the smart mobile phone, students can use the analog joystick and buttons on the gamepad to interact and communicate with the "Virtual Reality Assisted Teaching Software" in the smart mobile phone through the Bluetooth signal transmission. The complete system architecture is shown below as in Figure 1, and the actual wearing of VALID is shown below as in Figure 2.



VALID is a set of virtual reality teaching assisted software built by Unity game engine. Its main content is to assist students in exploring the glacier terrains. After the users finish wearing the VALID devices, they enter the first scene of the software, "VR Training Scene." Users at this stage are mainly to familiarize themselves with the rotation and movement of VR environment, and test whether the focus of VR Box glasses is in the right position and whether the gamepad functions properly. If students feel dizzy and uncomfortable at this stage, they are allowed to immediately stop the operation and take a rest. After the adjustment of hardware equipment is completed, the students can enter the second scene in the software, "Upper Glacier Terrain" to explore the terrains. There are four different terrain types in this scene for students to explore. After students complete the

exploration, they can pass through the transmission gate of the scene to switch to the next scene, "Downstream Glacier Terrain." After students complete the exploration, they can use the transmission gate of the scene to end the VALID system and remove their VR. Box glasses. When students leave the "Downstream Glacier Terrain Scene," VALID will automatically switch to the "Result Presentation" scene, and researchers will take the mobile phone out of the VR Box and record the terrains explored by students in VALID, as the data of task completion rates.

3.2. System scenes

In the part of system functions, this system divides the scene area in the software into four scenes: "VR Gamepad Movement Training," "Upstream Glacier Terrain," "Downstream Glacier Terrain" and "Result Presentation." These four scenes are introduced as follows in order:

3.2.1. VR gamepad movement training

The purpose of this scene design is to enable students to get familiar with the virtual reality environment after entering the VR learning environment. The scene herein is an empty room. After the equipment adjustment is completed and the students are familiar with the operation, they can leave the training scene by touching the gate on the wall with the gamepad. In this scene, students can first familiarize themselves with how to use the Bluetooth gamepad to move and fly, and adapt themselves to using the movement speed and flight mode of gamepad to assure that students can get the smoothest use experience.

3.2.2. Upstream glacier terrain scene

Glacial terrains can be divided into upstream and downstream glacier terrains. This system identifies the upstream and downstream terrain each with different scenes for students to learn. Students can switch and explore freely in between the two scenes. In both scenes, each has four terrain types requiring students to observe. After discussing with the experts, the upper glacier terrain is decided to focus on the four main terrain types: "Cirque Terrain," "Cirque Lake Terrain," "Horn Terrain "and "Arête Terrain" (Figure 3).



Figure 3. (a) Horn terrain; (b) Arête terrain; (c) Cirque lake terrain; (d) Cirque terrain

Students can move freely in the glacier terrains and look for the four glacier terrain types, and a "signpost" around each terrain type is placed to inform students that they are around the terrain type so as to enable students to find the correct observation target. If students are close to the signpost within a certain distance, the system will send out a bell sound to remind students that they reach the terrain type, and automatically record the

progress of students' exploration in the said terrain type. After the students finish the exploration, they can find a transmission gate in the scene. If the students approach and enter the transmission gate, the system will automatically switch the scene to the "Downstream Glacier terrain," realizing the effect of scene switching.

3.2.3. Downstream glacier terrain scene

As for downstream glacier terrain, we hope to focus the scene on four terrain types after discussing with the experts and geography teachers: "Fjord Terrain," "Glacial Trough Terrain," "Hanging Valley Terrain" and "Glacial Trough Lake Terrain" (Figure 4). The four main terrain types in the downstream glacier terrain scene all have their own signposts, allowing the students to interact with and explore the basic operation, similar to that of the upstream glacier terrain. After the students have finished the exploration, they can leave the learning scene and enter the "Result Presentation" scene through another gate marked with "Exit."

Figure 4. (a) Fjord terrain; (b) Glacial trough terrain; (c) Hanging valley terrain; (d) Glacial trough lake terrain



3.2.4. Result presentation scene

After students enter this scene from the "Downstream Glacier Terrain" scene, they can inform the instructor that they have completed the Geographic Assisted Learning Course. In the "Result Presentation" scene, it is not a VR mode, but a scene listing all the names of terrain types in the software and presenting them in red texts. If students used to access the terrain types in the software, the names of the terrain types listed in "Result Presentation" will be presented in green texts. By using this function, staff and teachers can know students' observation status in the software and whether students used to access the designated learning terrain types, whereby understanding and recording students' learning status, which are regarded as the data of "Task Completion Rate."

4. Research methods and experimental design

This study designed the following experimental procedures to explore whether senior high school students improve their learning effect for rare terrain through VALID (Figure 5). In this study, we invited two geography teacher to do the expert interview and also designed the exams and pre-scan the questionnaire items. In this study, two classes of senior high school sophomores were selected for comparison. There were 76 students in the two classes, 38 in the experimental group and the control group respectively. There were 10 boys and 28 girls in the control group where traditional teaching methods were given. The traditional teaching methods herein refer to the teaching media commonly seen in current senior high schools such as teaching through textbooks, PowerPoint, teachers' dictation and writing on blackboard. In the experimental group, there were 9 boys and 29

girls, giving additional VALID assisted learning to the traditional learning mode, and using VALID in class time to deepen their impression on the course content. The difference between the two learning modes was compared by the students' post-test scores.



The experiment lasted for two weeks, and there were two geography classes per week, wherein two groups of students used the traditional teaching mode to study the "New Zealand and Australia" unit of geography in the first week, while he students in the experimental group used VALID to assist learning in the second week. Teachers merged the two geography classes in the second week into one for a total of 100 minutes so that students can have sufficient time to use VALID. As shown in Figure 6, before the assisted class started, the teacher introduced the glacier terrains for 10 minutes, including all the glacier terrain types that appear in the software and the reasons for the formation of these glacier terrain types, so that the students can have a preliminary concept of the relevant knowledge of the assisted class.





After the introduction, the researcher started to introduce VALID, and introduce the process, operation mode and map objects through the computer version of "Virtual Reality Geographic Assisted Software" and projector. The researcher informed the students of the task objectives and matters needing attention in the process of using the software, and reminded them of immediately reporting any of their discomfort occurring in the process to the staff of the group and then immediately stopping using VALID.

After the researcher finished the introduction, he/she passed the space ability test books to the students. This spatial ability test book, of which the reliability is 0.82, comes from the "Spatial Ability Book" of the Academic Aptitude Test Book of the Ministry of Education. The test time was 20 minutes as specified in the instruction of

the "Spatial Ability Book." After all the students completed the space ability test book, the staff of each group started the assisted learning and asked students to wear the assisted teaching devices (Figure 7). The entire operation time of the assisted learning was about one hour as each student's use time was about five minutes, each group consisted of about 6-7 students and each student had to make individual hardware adjustment. After the students finished the operation, the staff passed the post-test questions and asked the students to start the test. There were 10 questions in the post-test designed by the high school teachers, the test scope of which was "New Zealand and Australia Unit." The time of the post-test was 20 minutes. After the completion of the test, the staff passed the questionnaires to collect the students' views and opinions on the Virtual Reality Assisted Teaching Course, and retrieved the questionnaires the next day. For the control group class, traditional teaching was given in the geography course of the second week, which was the same as that given in the first week, and a post-test was conducted before the end of the course. The test content was the same as that of the experimental group.



Figure 7. Using virtual reality assisted learning devices

5. Experimental results and discussion

In the experimental group, students were divided into "High Spatial Ability" and "Low Spatial Ability" groups through the spatial ability test. The "High Spatial Ability" students consisted of 17 students whose spatial ability scores were ranked as "Extremely high," "High" and "Ordinary." The "Low Spatial Ability" students consisted of 21 students whose spatial ability scores were ranked as "Low" and "Extremely Low." The learning effect and learning performance of these two group students after finishing the VALID assisted course were observed.

In addition to dividing the students through their spatial ability in this study, students were also divided into two groups according to their final semester grades in geography of last semester: "High Geographic Achievement "group and "Low Geographic Achievement" group. The high geographic achievement group consisted of 15 students whose final semester grades in geography were ranked in the top 40 percent of all students, while the low geographic achievement group consisted of 15 students whose final semester grades in geography were ranked in the bottom 40 percent of all students. Moreover, whether there was difference in the performance of students with different learning achievement after using VALID for learning was observed.

The semester final grades in geography of the students in two classes were obtained after permission from a senior high school in Taichung, and were further used as the data of "Learning Achievement in Geography" in this study. In the aspect of evaluating the "Learning Effect," 10 test questions designed by the school geography teachers were used as the test questions, on which the grades of the students in two classes were used as the students' Learning Effect. In the aspect of collecting the spatial ability data, this study adopted the "Spatial Relationship" test book of the "Academic Aptitude Test Scale of College Entrance Examination Center" to test students' spatial ability. Students' spatial ability was divided into "Spatial Visualization," "Spatial Perception" and "Spatial Positioning," on which the influence of different spatial ability on learning effect was analyzed in order.

In the following sub-sections, the experimental analysis of Sections 5.1, 5.2, and 5.3 are aimed to investigate the research question about "Understand the impact of learning achievement on learning performance"; In addition, the research question about "Understand the impact of spatial ability on learning performance" is discussed in Sections 5.4, 5.5, and 5.6. All experiments are designed and investigated for "Construct a VR-assisted learning system to improve students' learning method and increase learning effectiveness."

5.1. Comparison of the geographical learning achievement between the students among the experimental group and the control group

This study first confirmed whether the two groups of students have the same learning achievement for the geography course so as to avoid the biased results caused by excessive differences in learning achievement. In this study, "Independent sample *t*-test" was used to test the difference of the students' final semester grades between the two groups. The number of samples in the experimental group and the control group was 38 respectively, while the average final semester grade of the whole class in the experimental group was 68.34, and that in the control group was 67.45. After performing the "Independent sample *t*-test," we can see in Table 1 shown below that the p value of the difference between the two groups was "not significant." Therefore, we confirmed that there was no statistically significant difference between the two groups in geographical learning achievement, whereby the subsequent study can continue.

T 1 1 1 1 1 1 1			
Table I Independent com	nla t test of learning achieven	ant among the experimental	group and control group
<i>Tuble 1</i> . Independent sam	ibic <i>i</i> -iest of learning achieven		

	Average value (Standard deviation)		DOF	t value	р	Effect
	Experimental group $(N = 38)$	Traditional group ($N = 38$)				size (d)
Learning	68.34 (1.64)	67.45 (1.05)	74	0.383	.703	0.089
achievement						

5.2. Comparisons of post-test scores between the students among the experimental group and control group

In order to observe whether the students in the experimental group and the control group have different learning effect in geography course, the scores of the test questions designed by the senior high school geography teachers were used as the data of learning effect for the students in two groups. The average score of the experimental group was 55.53, while that of the control group was 47.63. Using the "Independent sample *t*-test" to test the difference between the experimental group and control group, we can see in Table 2 shown below that the significance of the difference in scores between the two groups was 0.017, indicating that there was a significant difference between the two groups in the post-test scores. From this result, we can know that the students in the experimental group were significantly better than those in the control group in terms of "Learning Effect."

Table 2. Independent sample *t*-test of learning effect in the experimental group and control group

	1 1		U U		U U	
	Average value (Sta	indard deviation)	DOF	t	р	Effect
	Experimental group	Traditional group		value		size (d)
	(<i>N</i> = 38)	(N = 38)				
Learning effect	55.53 (1.64)	47.63 (1.05)	74	2.448	$.017^{*}$	0.561
N. * 05						

Note. **p* < .05.

5.3. Comparisons of the post-test scores between the students in high and low geographic achievement groups among the experimental group

In order to understand the difference in learning effect of students with different geographical learning achievement after undertaking the virtual reality assisted learning course, in this study, the students of the experimental group were divided into two groups: "High Achievement" group (ranked in the top 40%) and "Low Achievement" group (ranked in the bottom 40%) according to the students' final semester grads in geography. The sample number of the high achievement group was 15, and the low achievement group was 15. The average score of the high achievement group was 60.67, while that of the low achievement group was 51.33. After going through the "Independent sample *t*-test," as shown in Table 3 below, we can see that the significance of difference between the two groups was 0.029, representing there was a significant difference in learning effect between the students in "High Achievement" group and "Low Achievement" group after undertaking the virtual reality assisted learning course. As a result, we concluded that students with better level in geography would perform better when they take virtual reality assisted learning course, and from which we can infer that this course was more suitable for the students with complete basic knowledge of geography.

Tuote et comparis	on of rearing encorection	een me men and ton ae		it groups in	and emperi	memaa group
	Average value (Sta	andard deviation)	DOF	t value	p	Effect size
	High achievement	Low achievement				(d)
	group $(N = 15)$	group ($N = 15$)				
Learning effect	60.67(1.64)	51.33(1.05)	28	-2.297	0.029^{*}	0.839
*						

Table 3. Comparison	of learning effect	between the high and low	achievement groups in the e	experimental group
---------------------	--------------------	--------------------------	-----------------------------	--------------------

Note. **p* < .05.

5.4. Comparisons of the post-test scores between the students with high spatial ability and the students with low spatial ability among the experimental group

In order to verify whether students with different "Spatial Ability" may have different learning performance after undertaking the virtual reality assisted geography learning course, the spatial ability score of each sample in the experimental group was obtained by asking the students in the experimental group to fill in the spatial ability test, and the students were further divided into two groups, the "High Spatial Ability" group and the "Low Spatial Ability" group, according to the definition in the test book. There were 17 students with "High Spatial Ability" scores (above 30 points of spatial ability) as shown in Figure 8, while there were 21 students with low spatial ability scores (below 30 points of spatial ability). The average score of students with high spatial ability in the post-test was 60.00, while the average score of students with low spatial ability was 51.90. After going through the "Independent sample *t*-test," the significance of the difference between the two groups was found to be 0.049 as shown below in Table 4, indicating that there was a significant difference between the data of the two groups. This result shows that students with high spatial ability would perform significantly better than those with low spatial ability in virtual reality assisted learning.



Table 4. Independent sample *t*-test of learning effect of the students with different spatial ability in virtual reality Average value (standard deviation) DOE = t value n = Effect size

	Average value (standard deviation)		DOF	t value	p	Effect size
	High spatial ability	Low spatial ability				(d)
	(<i>n</i> = 17)	(<i>n</i> = 17)				
Learning effect	60.00 (1.64)	51.90 (1.05)	36	-2.04	0.049^{*}	0.079
N. * 05						

Note. **p* < .05.

5.5. Verification of the correlation among space ability, task completion rate and post-test scores

As this study intended to explore whether the spatial ability of senior high school students would affect their task completion rates of the virtual reality geography assisted learning course. Therefore, the spatial ability scores, learning effect test scores and their task completion rates in the VALID system platform were utilized to verify

the correlation among them. In the system, there are eight glacier terrain types which need students to explore. The system records the students' exploring status in the software. After the completion of the assisted teaching, the statistics were made per eight glacier terrain types to show how many glacier terrain types the students explored in total. The more the students explored, the higher their task completion rates were, and the full score was 100 points. As shown below in the Table 4-5, after performing the correlation verification through SPSS, the statistical significance of the correlation between spatial ability and task completion rates was 0.661, and the Pearson correlation was -0.074, wherein these numerical values indicated that the correlation between spatial ability and task completion rates was not significant; therefore, we concluded that students' spatial ability did not affect their task completion rates in the VALID system platform.

In order to explore whether spatial ability would affect students' post-test scores, we verified the correlation between the experimental group students' spatial ability scores and their post-test scores through the statistics of task completion rates. As can be seen below in Table 5, the statistical significance of the correlation between students' spatial ability and post-test scores was 0.139, and the Pearson's correlation was 0.244, wherein these numerical values indicated that there was no significant correlation between the spatial ability and the post-test scores of the experimental group students; therefore, we concluded that spatial ability did not affect students' post-test scores in this experiment. However, the aforesaid conclusion contradicted the previous argument that "students with better spatial ability" would perform better with their virtual reality-assisted learning course; as a result, this experiment continued to verify the correlation for the three ability facets of spatial ability individually, with a view to finding out which spatial ability facet would affect students' learning effect.

Table 4. Pearson's correlation among spatial ability, task completion rate and learning effect

	Spatial ability	Task completion rate	Learning effect
Spatial ability	-		
Task completion rate	074	-	
Learning effect	.244	164	=

5.6. Influence of different spatial ability facets on post-test scores

The spatial ability test questions used in this experiment had three different test facets: spatial visualization, spatial perception and spatial positioning. These three facets tested students' three spatial abilities respectively. The test question of spatial visualization defined in the test book is the ability of visualizing abstract concepts and objects, and forming pictures in the mind; since this ability is the most difficult to define, nearly half of the questions were used for test in the test book. The test question prototype of spatial perceptual is defined as "Imagine yourself in a three-dimensional space, and you are still able to maintain a good sense of direction toward the goal after different orientation changes," while spatial positioning is defined as to mainly test whether students can find the designated coordinates and directions through instructions and conditions, and clearly locate the appropriate location. This study considered that the three spatial ability facets mentioned above would affect students' performance in the virtual reality assisted learning course, and the number of the three abilities in the test book was 19, 4 and 6, respectively. We calculated the scores of the students in the three facets respectively, whereby achieving the results of the students' three spatial abilities, and then observing whether different spatial ability would affect students' learning effect in virtual reality assisted learning geography course. Pearson correlation test was performed through SPSS in the following by using the experimental group students' grades on three spatial ability facets, their post-test scores and task completion rates to verify the correlation among students' spatial ability, learning effect and task completion rates.

5.6.1. Correlation among spatial visualization, post-test scores and task completion rates

Spatial visualization is one facet of spatial ability test. Pearson correlation test was performed through SPSS to verify the correlation among students' spatial visualization scores, post-test scores and task completion rates. In Table 6, the statistical significance of the correlation between students' spatial visualization ability and their post-test scores was 0.044, and the Pearson's correlation was 0.329, wherein these numerical values indicated that students' spatial visualization ability was positively correlated with their post-test scores, and that the better they performed well in the spatial visualization facet, the better their post-test scores were.

Table 0. Pearson's correlation between spatial visualization and learning effect			
	Spatial visualization	Task completion rate	Learning effect
Spatial visualization	-		
Task completion rate	176	-	
Learning effect	.329*	164	-
<i>Note.</i> $*p < .05$.			

CC 4

5.6.2. Statistical significance of the correlation among spatial positioning, post-test scores and task completion rates

Spatial positioning ability is the second facet of spatial ability test, the correlation verification of which was performed together with the students' post-test scores and task completion rates. As shown below in Table 7, we can see that the statistical significance of the correlation between students' spatial positioning ability and posttest scores was 0.402, and the Pearson's correlation was -0.140, while the statistical significance of the correlation between spatial positioning ability and task completion rates was 0.175, and the Pearson's correlation was 0.224, wherein these numerical values indicated that neither the correlation between students' spatial positioning ability and their post-test scores nor the correlation between students' spatial positioning ability and their task completion rates was significant; from which, we concluded that students' spatial positioning ability did not significantly affect their post-test scores in virtual reality assisted learning course and their learning performance in the software.

Table 7. Pearson's correlation between spatial positioning and learning effect

		1 0 0	
	Spatial positioning	Task completion rate	Learning effect
Spatial positioning	-		
Task completion rate	140	-	
Learning effect	.224	164	-

5.6.3. Statistical significance of the correlation among spatial perception, post-test scores and task completion rates

"Spatial Perception Ability" is the third facet of spatial ability test. After performing Pearson's correlation test, the statistical significance of the correlation between spatial perception ability and post-test was 0.325 and 0.733, respectively as shown below in Table 8, wherein there numerical values indicated that the students' spatial perception ability had no significant correlation with their post-test scores and task completion rates in this course; therefore, we concluded that the students' spatial perception ability did not affect their learning performance in the course. Through the analysis of three facets of spatial ability, we concluded that the facets other than "Spatial Visualization" did not affect the students' task completion rates in the software and their posttest scores, which may probably be due to the fact that there were more test questions in the facet of "Spatial Visualization," thus occupying a large proportion of scores in the whole spatial ability test, from which we therefore concluded the reason why there was a significant difference between the students' spatial ability and their post-test scores.

Table 8. Pearson's correlation between spatial perception and learning effect

		X	
	Spatial perception	Task completion rate	Learning effect
Spatial perception	-		
Task completion rate	057	-	
Learning effect	009	164	-

6. Conclusion

6.1. Research contribution

The conclusions of this study are summarized as the following.

6.1.1. Using virtual reality assisted learning devices to assist students in learning the "New Zealand and Australia" unit of geography course can effectively improve students' learning effect against that with traditional learning

Through the students' learning effect data obtained in the experiment and the verification via statistical analysis, we conclude that the students using VR devices have outstanding performance in the test of learning effect. In the past, many studies related to virtual reality also pointed out that using virtual reality technology to assist students in learning effectively improve their learning effect (Ebert & Tutschek 2019; Grivokostopoulou et al., 2016), while the results of this study can justify the previous literature.

In addition to deepening students' impression, using VR devices for learning is a brand-new teaching mode for students; therefore, it can attract students' attention and improve their learning motivation. In previous studies, it was also mentioned that the application of virtual reality can effectively enhance the learning motivation of adolescents (Vishwanath et al., 2017; Huang et al., 2016). The results of this study allow us to reasonably infer that experimental group students have better learning effect.

6.1.2. Students with better geographical achievement have better learning effect in virtual reality assisted learning course

The results of this study indicate that this course is of great help to students with solid knowledge of geography. Because the high achievement group students originally had perfect knowledge of glacier terrain and further verified it through virtual reality assisted learning devices, the high achievement group students would therefore perform even better. On the contrary, students with weak knowledge of geography had difficulty in verifying their own geographical concepts even in virtual environments because of their originally incomplete knowledge concepts; therefore, it is difficult for them to obtain learning effect from virtual reality assisted learning course. Based on this conclusion, we suggest that students enrich themselves with enough relevant knowledge (Sullins et al., 2018; Lai et al., 2021) before carrying out this course, and that teachers allow students to verify their knowledge by use of the features of virtual reality so that more learning benefits are brought to this course.

6.1.3. Students with better spatial ability have better learning effect in virtual reality assisted learning course

In this study, the "Spatial Ability Test Book of College Entrance Examination Center" was used to test the students' spatial ability. This test book divides spatial ability into three facets: "Spatial Visualization," "Spatial Perception" and "Spatial Positioning." Through the test on these three facets, we divided students into two groups: "High Spatial Ability" group (students with extremely high, high or ordinary spatial ability), and "Low Spatial Ability" group (students with low or extremely low spatial ability). By comparing the learning effect of the two groups, we find that students with high spatial ability would perform better in learning effect.

From the previous literature, we can know that spatial ability is a very important factor for virtual reality learning (Tchoubar et al., 2018); based on this result, we know that students with better spatial ability can get into the swing of the VR glacier terrain software more quickly, and have more time to observe different glacier terrain types. On the contrary, students with relatively poor spatial ability have to spend more time familiarizing themselves with VR environments, leading to the time compression of observing the glacier terrain types. As a result, students with better spatial ability would get better learning experience, the result of which was reflected in the test of learning effect.

6.1.4. There is a positive correlation between students' spatial visualization in spatial ability and their learning effect

After statistical analysis, it is found that "Spatial Visualization" in spatial ability positively would affect students' learning effect, but neither "Spatial Positioning" nor "Spatial Perception" had a significant correlation with students' geographical learning effect. Based on this result, we conclude that students with better "Spatial Visualization" would have better learning performance (Stieff & Uttal, 2015).

6.2. Research limitations

As this study required a mobile device to work with the VR Box to perform virtual reality imaging, the quality of the mobile device directly affected the user's experience. In addition, the fine game scenes and pictures not only burdened the memory of the mobile phone, but also consumed the battery rapidly. Not being able to be charged during the course, the system was designed not to provide students with the best resolution during the system design process.

In many studies, gender difference is a frequently discussed factor in exploring the learning effect and spatial ability of virtual reality. However, because the samples of this study were from the liberal arts students of a senior high schools, and the ratio of male students to female students was quite uneven, it was therefore difficult to obtain enough male and female samples for sample analysis.

In the course of this study, a total of 12 students expressed their dizziness. Some studies once pointed out that the use of immersive virtual reality should not be too long, five to ten minutes would be the best time for use, and the varying degrees of dizziness might occur based on individual differences. Based on the above reasons, the system was designed to set the completion time within five to seven minutes (Bouchard et al., 2011) in the system design process, trying to minimize the occurrence of dizziness among students as much as possible. The phenomenon of dizziness of human body greatly limits the teaching time and content of virtual reality, which is also a major issue still under discussion.

6.3. Future works

As this study focus on learning effect, spatial ability and learning achievement, topic of students' experience were less explored, such as whether students can achieve their "Experience of Flow" in the process of virtual reality assisted learning and be absorbed in the course, and comparing students' "Learning Motivation" in traditional learning and with that in virtual reality-assisted learning, as well as the discussion on students' internal and external motivation, all of which are the parts for future study. Moreover, it would be more effective to improve the learning mode and system and provide students with an even better learning environment through the survey of satisfaction and motivation.

Acknowledgement

The authors wish to thank the Ministry of Science and Technology of the Republic of China for financially supporting this research under Contract No. MOST-110-2628-H-005-002 and MOST-108-2511-H-005-001-MY3.

References

Bashabsheh, A. K., Alzoubi, H. H., & Ali, M. Z. (2019). The Application of virtual reality technology in architectural pedagogy for building constructions. *Alexandria Engineering Journal*, 58(2), 713-723.

Bouchard, S., Robillard, G., Renaud, P., & Bernier, F. (2011). Exploring new dimensions in the assessment of virtual reality induced side effects. *Journal of Computer and Information Technology*, 1(3), 20-32.

Carvajal, D. A. L., Morita, M. M., & Bilmes, G. M. (2020). Virtual museums. Captured reality and 3D modeling. *Journal of Cultural Heritage*, 45, 234-239.

Chen, C. C., Hsiao, K. L., & Li, W. C. (2020). Exploring the determinants of usage continuance willingness for locationbased apps: A Case study of bicycle-based exercise apps. *Journal of Retailing and Consumer Services*, 55, 102097. https://doi.org/10.1016/j.jretconser.2020.102097

Chen, C.C., & Huang P. H. (2020). The Effects of STEAM-based mobile learning on learning achievement and cognitive load. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2020.1761838

Chen, M., Lin, H., & Lu, G. (2017). Virtual geographic environments. In Richardson, D., Castree, N., Goodchild, M. F., Kobayashi, A., Liu, W., & Marston, R. A. (Eds.), *International Encyclopedia of Geography: People, the Earth, Environment and Technology* (pp. 1-11). Wiley-Blackwell

Donnon, T., DesCôteaux, J. G., & Violato, C. (2005). Impact of cognitive imaging and sex differences on the development of laparoscopic suturing skills. *Canadian Journal of Surgery*, 48(5), 387–393.

Ebert, J., & Tutschek, B. (2019). Virtual reality objects improve learning efficiency and retention of diagnostic ability in fetal ultrasound. *Ultrasound in Obstetrics & Gynecology*, 53(4), 525-528.

Grivokostopoulou, F., Perikos, I., & Hatzilygeroudis, I. (2016). An Innovative educational environment based on virtual reality and gamification for learning search algorithms. In 2016 IEEE Eighth International Conference on Technology for Education (pp. 110-115). https://doi.org/10.1109/T4E.2016.029

Hsu, T. C., & Chan, C. Y. (2018). The Effects of applying virtual-reality implementation for reflection after contextual mobile learning. In 26th International Conference on Computers in Education, ICCE 2018 (pp. 752-755). Asia-Pacific Society for Computers in Education.

Hsiao, K. L., Lytras, M. D., & Chen, C. C. (2020). An In-app purchase framework for location-based AR games: The Case of Pokémon Go. *Library Hi Tech*, *38*(3), 638-653.

Huang, H. M., Liaw, S. S., & Lai, C. M. (2016). Exploring learner acceptance of the use of virtual reality in medical education: A Case study of desktop and projection-based display systems. *Interactive Learning Environments*, 24(1), 3-19. https://doi.org/10.1080/10494820.2013.817436

International Data Corporation (2020). Worldwide augmented and virtual reality spending guide. IDC Press.

Ip, H. H., Wong, S. W., Chan, D. F., Byrne, J., Li, C., Yuan, V. S., & Wong, J. Y. (2018). Enhance emotional and social adaptation skills for children with autism spectrum disorder: A Virtual reality enabled approach. *Computers & Education*, *117*, 1-15. https://doi.org/10.1016/j.compedu.2017.09.010

Lai, C. F., Zhong, H. X., & Chiu, P. S. (2021). Investigating the impact of a flipped programming course using the DT-CDIO approach. *Computers & Education*, 173, 104287. https://doi.org/10.1016/j.compedu.2021.104287

Lee, E. A. L., & Wong, K. W. (2014). Learning with desktop virtual reality: Low spatial ability learners are more positively affected. *Computers & Education*, 79, 49-58.

Lemke, K. A., & Ritter, M. E. (2000). Virtual geographies and the use of the internet for learning and teaching geography in higher education. *Journal of Geography in Higher Education*, 24(1), 87-91.

Limniou, M., Roberts, D., & Papadopoulos, N. (2008). Full immersive virtual environment CAVETM in chemistry education. *Computers & Education*, 51(2), 584-593.

Lin, H., Chen, M., & Lu, G. (2013). Virtual geographic environment: A Workspace for computer-aided geographic experiments. *Annals of the Association of American Geographers*, 103(3), 465-482.

Lin, J. H. T. (2017). Fear in virtual reality (VR): Fear elements, coping reactions, immediate and next-day fright responses toward a survival horror zombie virtual reality game. *Computers in Human Behavior*, 72, 350-361.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A Metaanalysis. *Child Development*, 56(6), 1479-1498.

Lovreglio, R., Gonzalez, V., Amor, R., Spearpoint, M., Thomas, J., Trotter, M., & Sacks, R. (2017). The Need for enhancing earthquake evacuee safety by using virtual reality serious games. In *Lean & Computing in Construction Congress (LC3 2017)* (pp. 1-9).

Lv, Z., & Li, X. (2015). Virtual reality assistant technology for learning primary geography. In Z. Gong, D. Chiu, & D. Zou (Eds.), *Current Developments in Web Based Learning. ICWL 2015. Lecture Notes in Computer Science* (Vol. 9584, pp. 31-40). Springer, Cham. https://doi.org/10.1007/978-3-319-32865-2_4

Merchant, Z., Goetz, E. T., Keeney-Kennicutt, W., Cifuentes, L., Kwok, O. M., & Davis, T. J. (2013). Exploring 3-D virtual reality technology for spatial ability and chemistry achievement. *Journal of Computer Assisted Learning*, 29(6), 579-590.

Metoyer, S. K., Bednarz S. W., & Bednarz R. S. (2015). Spatial thinking in education: Concepts, development, and assessment. In Muñiz Solari, O., Demirci, A., & Schee, J. (Eds.), *Geospatial Technologies and Geography Education in a Changing World. Advances in Geographical and Environmental Sciences* (pp. 21-33). Springer. https://doi.org/10.1007/978-4-431-55519-3_3

Roch, P. J., Rangnick, H. M., Brzoska, J. A., Benner, L., Kowalewski, K. F., Müller, P. C., Kenngott, H. G., Müller, B. P., & Nickel, F. (2018). Impact of visual-spatial ability on laparoscopic camera navigation training. *Surgical Endoscopy*, *32*(3), 1174-1183.

Stieff, M., & Uttal, D. (2015). How much can spatial training improve STEM achievement? *Educational Psychology Review*, 27(4), 607-615.

Sullins, J., Acuff, S., Neely, D., & Hu, X. (2018). When knowledge isn't power: The Influence of prior knowledge on question generation training. *Journal of Educational Multimedia and Hypermedia*, 27(2), 245-265.

Tascón, L., Castillo, J., León, I., & Cimadevilla, J. M. (2018). Walking and non-walking space in an equivalent virtual reality task: Sexual dimorphism and aging decline of spatial abilities. *Behavioural Brain Research*, 347, 201-208.

Tchoubar, T., Sexton, T. R., & Scarlatos, L. L. (2018). Role of digital fluency and spatial ability in student experience of online learning environments. In Arai, K., Kapoor, S., & Bhatia, R. (Eds.), *Intelligent Computing. SAI 2018. Advances in Intelligent Systems and Computing* (Vol. 857, pp. 251-264). Springer, Cham. https://doi.org/10.1007/978-3-030-01177-2_18

Vishwanath, A., Kam, M., & Kumar, N. (2017). Examining low-cost virtual reality for learning in low-resource environments. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (pp. 1277-1281). https://doi.org/10.1145/3064663.3064696

Walczak, S., & Taylor, N. G. (2018). Geography learning in primary school: Comparing face-to-face versus tablet-based instruction methods. *Computers & Education*, 117, 188-198.

Woodford, C. (2019). What is virtual reality. Explain that stuff. https://www.explainthatstuff.com/virtualreality.html

A Grounded Theory Study of the Psychological Distance in Online Education

Ling Zhang¹, Jiang Ye^{2*} and Jing-xuan Wang¹

¹College of Management Science, Chengdu University of Technology, China // ²College of Earth Sciences, Chengdu University of Technology, China // zhangling@cdut.edu.cn // yejiang@cdut.edu.cn //

wjx7828@163.com

*Corresponding author

ABSTRACT: In educational reform, people actively promote educational innovation by applying intelligent technology. As the main participants in education, people experience a series of psychological and cognitive changes in the teaching process. This autonomy and uncertainty will directly affect the effect of network teaching. Due to the variability in the characteristics of individuals, following the psychological perception and subjective value of people and optimizing the development of wisdom education with a reasonable technical cognitive attitude have become a focus. This study uses the grounded theory method to conduct in-depth interviews and questionnaire surveys on 330 e-learners and extracts 400 codes, 38 concepts, 9 categories and 4 core categories. This paper reconstructs the theoretical model of psychological distance in the network education process and identifies four new dimensions: cost distance, initiative distance, control distance and interaction distance. In addition, the four dimensions reflect four core value needs of learners for online education: convenience, self-efficacy, self-identity and binding force. The paper provides a good theoretical basis to improve the user experience and satisfaction in online education and optimize the level of intelligent education.

Keywords: Online education, Psychological distance, Grounded theory

1. Introduction

The sudden onset of COVID-19 has challenged education systems, and education reform, which has been pushed to the forefront, cannot be delayed. All courses have been transferred to online, which is a real and thorough internet + teaching practice for all universities. How to better apply smarter technology in education and provide informative and intelligent education has become a new focus of scholars. Overall, the development and progress of online education practices have occurred with the progress and use of educational technology. Technology enables educators to overcome common learning barriers in online education. For example, Cantabella et al. (2020) proposed a new tool to evaluate students' satisfaction using smileys, and Villagrá-Arnedo et al. (2020) proposed the construction of an achievement prediction system to help teachers gain insight into students' learning trends. However, although technology has done its best to bring online learners "closer" to schools and courses, the impact of "distance" in online education is still far-reaching. This is because as the main participants in education, the psychological perception of people is changing and uncertain. In the distance learning process, what types of participation in relevant cognitive and psychological processes are best (Breves & Schramm, 2021)? How can we highlight the proximity of learning (Afrouz & Crisp, 2021)? We need to put forward the right cognitive attitude toward technology. We must pay attention to learners' user experience from people's psychological perception and subjective value and improve the effect of network teaching through user satisfaction to optimize the development of intelligent education.

1.1. Literature review

1.1.1. Research on the user experience of online education

Some scholars have used the theory of "sense of social presence" to analyze user experience problems and sought to enhance the effect of online education. For example, Kwon (2011) explored the impact of social presence on learning outcomes and the relationship between social presence and learners' characteristics. Kožuh et al. (2015) analyzed the influence of social presence and interaction on the success of students in the learning environment. However, teaching scene simulation is only one form, and not every learner can actively participate in consciousness and achieve good learning results. Many scholars have also researched the influence of other factors. By testing a model of online learner participation, Vayre and Vonthron (2017) demonstrated that self-efficacy only mediates between the sense of community and the sense of participation. Bellotti et al. (2013) and Orlando (2015) discussed using game design in online education to keep students interested. Warren and Nash (2019) explained how online consulting education used the art of expression to increase people's awareness.

Farrell et al. (2016) mentioned that the interaction between learners and online interfaces, content, and other learners was an important consideration in designing online education courses. Matcha et al. (2019) created a model for user-centered learning analytics systems.

In the research on online education user experience, scholars observe different perspectives (such as selfefficacy, transaction distance and satisfaction, interest stimulation, perceived service quality, and emotional interaction), which resulted in different attribution analyses of online education user experience. Both quality of user experience and satisfaction are the results of multiple psychological perceptions of the learners. Can we explore the advantages and disadvantages of user experience and identify the factors that affect the user satisfaction from multiple angles and distances of learners' psychological perception?

1.1.2. Research on psychological distance

The concept of psychological distance was proposed by Bullough (1912) in the field of philosophy and aesthetics in 1912. He pointed out that the generation of beauty arises from the psychological distance between the subjective perception of the observer and the artwork. Liberman and Trope (1998) first introduced psychological distance into the field of social psychology in 1998. They suggested that psychological distance refers to people's subjective experience of approaching or moving away from the reference point of a certain thing and that they make judgments based on how they feel in the moment (Trope & Liberman, 2003). Dhar and Kim (2007) asserted that psychological distance is the subjective distance between the parties and the event in the psychological space of the parties. Therefore, the definition of psychological distance mainly involves three points: first, an individual has a subjective experience of the target; second, the experience has different psychological dimensions; and third, the origin of the distance is the observer itself. Integrating the thoughts of several researchers, this study summarizes the concept of psychological distance as follows: psychological distance between the self and the event in the psychological space of the subjective distance between the self and the event in the psychological space of the self when the self is taken as the origin of the distance.

Tolman (1932) was the first to define the dimension of psychological distance. Later, Bar-Anan et al. (2006) explored the correlation between construal level theory (CLT) (people's mental representation of cognitive objects or events shows different degrees of abstraction (Liberman & Trope, 1998; Trope & Liberman, 2003) and psychological distance. Amit et al. (2009) and Bar-Anan et al. (2006) asserted that there is a potential, automatic and regular relationship between psychological distance and CLT. Bar-Anan et al. (2007) proposed four dimensions of psychological distance based on CLT. Most modern studies are based on Trope's four psychological distances:

- Temporal distance: the distance between the target event in the past or the future and the individual in time;
- Spatial distance: how far the stimulus or target event is from the center in space;
- Social distance: the affinity or similarity between the social object and center;
- Hypothetical distance: the proximity of an event or object to reality or the probability of its occurrence or existence.

As the distance system of CLT, the psychological distance is not perfect. Whether there are other distance dimensions has always been controversial. Liberman et al. (2007) claimed that in addition to the above four dimensions, the psychological distance should include others. Based on the perspective of social psychology, this paper reviews the relevant literature and summarizes the psychological distance dimensions studied by many scholars (see Table 1).

Researchers	Dimensions
Tolman (1932)	Spatial dimension, temporal dimension, energy input dimension
Engebretson (1973)	Interaction distance
Boroditsky & Ramscar (2002)	Temporal dimension, spatial dimension
Bar-Anan et al. (2006)	Temporal distance, social distance, spatial distance, probability of
	occurrence
Liberman et al. (2007)	Spatial dimension, temporal dimension, social dimension, hypothetical
	dimension
Fiedler (2007)	In addition to the four dimensions of psychological distance proposed in
	traditional studies, information distance, perspective distance, emotional
	distance and experience distance can also be used as research dimensions

Table 1. List of psychological distance research dimensions

Maglio & Polman (2014)	Based on the four classical dimensions, the research on psychological distance is expanded from a static spatial position to a dynamic moving point in space
Chen & Guibing (2014)	Temporal distance, social distance, probability distance
Chen & Li (2018)	Experiential distance, behavioral distance, emotional distance, cognitive
	distance, spatial and temporal distance, objective social distance
Horvath (2018)	Identification distance, safety distance, value distance, control distance
Li et al. (2019)	Cognitive distance, emotional distance, expectancy distance, behavioral
	distance
Liu et al. (2020)	Spatial distance, temporal distance, social distance, hypothetical distance

1.1.3. Research on psychological distance in the field of education

In recent years, the theories of CLT and psychological distance have also been widely used in the field of education. Vaughn and Baker (2004) found that psychological distance contributes to teaching effects and satisfaction. Lee (2010) examined the relationships among the psychological distance perceptions of students and online teachers, academic performance, and willingness to continue online learning. Zhbanova and Rule (2014), using CLT, noted that focusing on distant (distal), in contrast to near (proximal), content promotes mental levels of abstract thought increasing creative performance. Ho et al. (2015) discussed the problem that users' CLT and perception of e-learning systems affect their willingness to adopt. Neroni et al. (2015) explored the biological and psychological factors associated with learning achievement in adult distance education. Lee et al. (2017) took four university courses as the object of study to explore the effect of cognition of instructors, tutors and students on the role of tutoring and whether tutoring affects the psychological distance between different types of participants. Sungur et al. (2017) discussed the meaning of psychological distance and CLT in the context of online inspiration and persuasion. Weidlich et al. (2018) examined the relationship between transaction distance and satisfaction in the context of distance education.

The literature reveals that scholars have conducted extensive research on students' behavior, attitude and views from the perspective of psychology, enhanced their comprehensive understanding of education, and laid a good foundation for the development of network education research. However, the theory and framework of psychology have not matured. According to CLT, the underlying psychological mechanism is not fully developed, including the psychological distance. We found that in the new environment of online education, the four dimensions widely used by most scholars (Bar-Anan et al., 2007) cannot explain the psychological perception and behavioral results of online learners well. For example, a student has strong expectations for the learning effects of an online course and has a well-functioning computer and independent space, but he often cannot complete the course and always gives up halfway through it. Interviewing different respondents, we found that the reasons for this occurrence are diverse: because there are too many learning resources, they develop choice anxiety and linger in multiple courses; although they have a clear goal, they cannot successfully complete the course due to laziness or procrastination; their learning interest is not stimulated due to the lack of full interaction with teachers or synchronous learners, whether in troubleshooting or creating classroom atmosphere. Are there other new dimensions of psychological distance that can explain the user experience and behavior of learners in online education? This study will construct a new dimension of psychological distance in the online education process based on the psychological perception of online learning users and the extensive collection of all types of learner data.

1.2. Grounded theory

Grounded theory was first proposed by two American scholars, Glaser and Strauss (1967) in The Discovery of Grounded Theory. Grounded theory is not a fixed theory but a research approach or "methodology" in the field of qualitative research. The purpose of the research is to put forward the theoretical concept and clarify it through the systematic collection and analysis of empirical data, existing literature and researchers' knowledge to excavate the connotation and extension of the concept from practice. If the connotation and extension of theoretical concepts have been well explained and widely supported, quantitative research is suitable for verification; if the connotation and extension are unclear or controversial theoretical concepts remain, it is more suitable to adopt qualitative research, especially qualitative research based on the grounded theory research method. Grounded theory is especially suitable for research fields that lack theoretical explanation or the explanatory power of existing theories and research on the micro- and action-oriented social interaction process (Corbin & Strauss, 2014). However, in grounded theory, discovery is guided by data collection and inductive analysis instead of by existing theoretical models (Sahoo et al., 2015). It is innovative instead of verifiable.

Some scholars have also proposed that content analysis and coword analysis can be used to construct some theories, but the difference between them and grounded theory is that grounded theory is the overall methodology to systematically collect and analyze data. Content analysis is the collection and analysis of specific data in specific situations and only one of many data collection methods based on grounded theory. Coword analysis is only a part of content analysis, such as the relationship between word frequency and word meaning, but this simple relationship cannot fully satisfy the requirements of using grounded theory to explain the real situation (Krippendorff, 2018).

This study focuses on the issue of psychological distance, which has been controversial among scholars. In addition, in the network education environment, the real experience and behavior of learners can no longer be explained by the existing dimension of psychological distance. Therefore, the grounded theory method is appropriate in this study.

A literature review has found that this method is rarely used in the psychological distance in the online education process. Only Green (2013) used the grounded theory method to describe the social process of relative distance, and it was found that participants paid more attention to some learning achievements than others. However, it does not truly involve the discussion of the complete dimension of psychological distance in the online education process. Therefore, this study completely relies on learner data, uses grounded theory to summarize and refine the dimension of psychological distance, and seeks to build a theoretical model of the psychological distance against the background of network education.

2. Materials and methods

2.1. Research steps

We follow the Glaserian or "classic" method to generate the grounded theory. This approach begins with identifying categories of behavior (open coding) to determine the core categories that represent the central idea or structure of the research. Then, selective classification analysis of the concepts and core categories is conducted (selective coding). Finally, theoretical coding is used to generate concepts that explain the relationships between the integrated core categories and other elements (Gasson & Waters, 2013).

The specific research steps of this paper are as follows: first, user groups are selected according to the theoretical sampling requirements; second, secondary information is collected from official online education websites, official apps, major forums, and other channels; third, according to the secondary data collation results, an interview outline is drawn up, and in-depth interviews are conducted to obtain primary data; fourth, based on the rules of grounded theory and coding techniques, the theoretical model of the new dimension of psychological distance in online education is extracted; and fifth, a saturation test is conducted.

2.2. Sampling procedure

American online education is mainly supplied by large-scale colleges, public colleges, and private nonprofit colleges and shows high concentration. According to a report released by the Sloan Alliance in 2017, 47.7% of online education students are concentrated in approximately 5 percent of schools. The top 47 schools, accounting for only 1% of the total number of schools, contain 23% of the total online education population. In China, according to a 2018 white paper on China's online education users, accounting for 62.1% of the total number of online education users; learners aged 16 to 45 account for 84%.

Young people, especially those still in the stages of education and career advancement, and middle-aged people constitute the main body of online education users.

Consequently, high school students, college students, graduate students, and young people entering the workforce were selected as the respondents in this study. The age range was concentrated in the 17-35 age group, with an equal number of males and females, and we followed the nonprobability sampling principles of purposeful sampling and heterogeneous sampling. To ensure that the interviewees could provide the information needed for this research, we sought to cover as many different types of interviewees as possible, recruiting participants from different industries, with different education levels, with different learning objectives and from

different regions. When using a purposeful sampling method, potential participants are deemed eligible, and their insights enable exploration of the phenomena of interest (Creswell, 2006).

2.3. Sources

From January to April 2020 and January to March 2021, we collected and sorted the primary and secondary data, providing opportunities for data testing through a variety of data collection techniques (Holt et al., 2017). Secondary information was mainly collected from the official websites, official apps, and official microblogs of MOOC (massive open online courses) and NetEase Cloud classrooms. Some data were also obtained from ZhiHu.com. Through keyword searches, we obtained a total of 35 relevant materials. The search content was mainly related to online education and learning environments, learning efficiency, classroom interaction, personalization, learning self-discipline, Q&A and other high-frequency keywords.

The primary data were collected through in-depth interviews and questionnaires. According to Patton (2014), creative fieldwork means using all of the researcher's senses to experience and understand what is happening. Creative insight comes from being directly involved in the situation under study. Therefore, the focus of this study was on collecting original data in the form of in-depth interviews. Before the interviews, an in-depth interview outline was designed to develop the research in accordance with the results of the secondary data collection. During the interview process, the focus of and questions asked during the interviews did not strictly follow the outline but were adjusted according to the interviewees and the progress of the interview. It is not possible to know the exact number of study participants who will be sampled before a study begins (Foley & Timonen, 2015). Ultimately, through screening, we conducted in-depth interviews with 52 respondents online.

The questionnaire method in this study involved integrating the interview outline into a set of items and sending them to the research subjects via the internet to supplement or verify the interview results. A total of 278 supplementary questionnaires were analyzed.

For convenience, the in-depth interviews and questionnaire surveys were mainly conducted in China. The respondents were 330 online education users in Sichuan, Beijing, Jilin, Qinghai, Shanxi, Shanghai, Shenzhen, Yunnan and Guangzhou, and most were high school students, college students, and graduate students, with a few young men who had just started working. The interviewees were interviewed for approximately half an hour. The interview steps were as follows:

2.3.1. Interviewee recruitment

In general, grounded theory research involves deliberately selecting participants who can provide valuable insight into the research topic (Sbaraini et al., 2011). Therefore, we used the following two approaches. (i) The online education platform "course evaluation" or "discussion board" features were used to find interviewees, explain the purpose of the research, ensure understanding and obtain consent, and conduct the interviewes. (ii) Following the theoretical sampling principle, this study selected various types of interviewees from different regions, identities, age stages, and industries, as shown in Table 2.

Table 2. Data collection method					
	High school	College	College students just	A graduate	Combined
	students	students	entering the workplace	student	(person.)
	(person.)	(person.)	(person.)	(person.)	
Questionnaires	84	102	47	45	278
In-depth interviews	12	24	9	7	52
Combined	96	126	56	52	330

2.3.2. Design of the interview content

The interview outline used in this study was based on the prior collection and analysis of secondary data. The questions related to three stages of online education.

(i) Before online education (preparation)

- What are the reasons you chose online education? Which platform...
- What's your expectation for every online learning course? What is the basis?

(ii) Online education in progress (actual results)

- Offline learning status? Online learning status and mindset? What makes these differences for you?
- Which learning efficiency do you think is more efficient for you? Why?
- Will your psychological feelings affect your final learning effect? If so, what is your specific psychological feeling?

(iii) After online education (experience)

- Usually, after your actual experience, have you met your expectations for online learning? If not, what do you think is the reason?
- Do you take the initiative to give feedback with learning questions, feelings and suggestions to teachers and platforms? What is the reason?

2.2.3. Interview principles

Lofland (2006) suggested that the principal obligation of the researcher to the respondents is to guarantee anonymity through a "confidentiality agreement." Therefore, the respondents' responses were kept strictly confidential in this study. The interview content was saved in the form of text. The data were sorted within 24 hours after each interview to ensure that the meaning expressed by the interviewes was accurately recorded.

3. Results

After the data were collected and preliminarily sorted, grounded theory was applied to conduct further in-depth analysis in three main steps: open coding, axial coding and selective coding (Corbin & Strauss, 1990). The grounded theory terms relevant to data analysis are explained below (Corbin & Strauss, 2014).

- Concept: A conceptual label attached to an individual event, case, or phenomenon.
- Category: When a group of concepts all refer to the same phenomenon, they are identified by a higher and more abstract concept called a category.
- Core category: A categorical concept used to encompass the events or phenomena reflected in the case as a whole.
- Dimension: A subdivision of a category based on certain attributes that enables a better understanding of the meaning of the category.
- Context: If the case has clear category dimensions, this case-specific content can be referred to as the location of the dimension. The context is a set of special conditions in which action or interaction strategies occur.
- Storyline: The main thread of the case story summarized in one sentence.

3.1. Open coding

Open coding refers to encoding the similarities and differences in the data; the data are constantly compared with new data indices and concepts to create new concepts (Patton, 2014). In this process, the names of the concepts and categories come from the literature, interview records, and discussion among the researchers. The name can be a word, a phrase, or even a short sentence, not just a simple "abstract." In this study, Michael Quinn Patton's mutual comparison method was used to determine whether to continue data collection. If 20 consecutive texts were compared with each other and no new content was found, data collection was stopped. Examples of concepts and open coding from this study are presented in Table 3. The code ai represents the original interview sentence. The code Ai indicates a refinement of the original interview statement. The code bi refers to the conceptual content. The code Bi represents the categorized content.

Tuble 5. Examples of concepts extracted field	mi the original data (soried	by source)
Interview notes (ai)	Memo (ai)	Conceptualization (bi)
(a2) At noon at the company lunch, I can also watch some micro courses on software application skills on my mobile phone.	A2 Not limited by learning places	b9 Study anywhere
(a29) Usually, I can easily find the online courses I need, but sometimes I can find too many courses. It may be difficult to choose the ones that truly suit me.	A29 Find suitable online courses online	b17 Accuracy of resource selection
(a32) In my studies, if I constantly encounter problems, I will be very flustered and depressed. If I know that the people studying with me are the same, I will feel better and have more confidence to solve problems.	A32 Common experiences of CO learners	b33 Peer motivation
(a147) For simple or mastered knowledge content, I can choose to watch it faster to save learning time and improve efficiency.	A147 Watch course at double speed	b6 Save time watching learning
 (a266) I cannot concentrate on my studies alone. I always have to think about other irrelevant things in my mind.	 A266 Inability to concentrate	 b26 Short attention span
(a384) This teacher is a famous teacher at Peking University. I think I can learn better!	A384 Excellent teachers enhance learning confidence	b23 Expectations of teachers
•••		•••

Table 3. Examples of concepts extracted from the original data (sorted by source)

Through the open coding of the psychological distance data from the interviews on online education, 38 concepts were abstracted, as shown in Table 4.

3.2. Axial coding

Through the above collation and analysis of the data, we sought to establish preliminary relationships among the categories and concepts. By exploring these relationships, the concepts could be used to develop the categories in more detail (Foley & Timonen, 2015). Therefore, this paper now turns to the relationships among the categories.

The main task of axial decoding is to identify and establish the relationships among categories, which can be causal, similarity, difference, equivalence, structural, or functional relationships, among others.

The typical model is an important analytical tool in the grounded theory approach, and it is an effective means to connect and further explore the categories. It contains six aspects – causal conditions, phenomena, context, mediating conditions, action/interactive strategy, and results – which are used to guide the sorting and analysis of the categories (Corbin & Strauss, 1990). In the third edition of Strauss's book, Corbin modified this model by adding "emotion" to the "action/interaction" section (Corbin & Strauss, 2014).

With the help of this model, this paper constructs the relationships among the concepts and categories, as shown in Table 4.

Number	Category	Concept
B1	Time	b1 Commissioning time
		b2 Time for complete mastery of knowledge
		b3 Dressing time
		b4 Arrange time independently
		b5 Time to deal with interference events
		b6 Save time watching learning
		b7 Fragmented time utilization

Table 4. Correspondence between concepts and categories

B2	Space	b8 Learning environment is more comfortable	
		b) Shorton the distance from high quality curriculum resources	
B 3	Pasourcas	b10 Shorten me distance from high-quanty currentum resources	
D 5	Resources	b12 Multiple learning platforms	
		b12 Frae resources	
		b14 High quality resources	
		b15 Online knowledge is highly refined	
		b16 Online course can be reviewed repeatedly	
		h17 Accuracy of resource selection	
B4	Cognition	h18 Awareness of social needs	
21	Cogintion	b19 Cognition of self-knowledge level	
		b20 Knowledge of specific courses	
		b21 Self-motivated to learn	
B5	Anticipation	b22 Expectations of course content	
	1	b23 Expectations of teachers	
		b24 Expectations of learning outcomes	
		b25 Expectations about the learning process	
B6	Self-control	b26 Short attention span	
		b27 Poor ability to resist distraction	
		b28 Learning enthusiasm decreases gradually	
B7	Supervision	b29 Examination supervision	
		b30 Drive of the atmosphere	
		b31 Curriculum planning	
B8	Emotions	b32 A preference for courses	
		b33 Peer motivation	
		b34 Preference regarding teachers and platforms	
		b35 Strong learning atmosphere	
B9	Behavior	b36 Communication between teachers and students	
		b37 Peer discussion	
		b38 Classroom activity	

3.3. Selective coding

The aim of selective coding is to identify a core category that encompasses the other categories or supersedes them in terms of explanatory importance. The relations among categories constitute entity theory (Foley & Timonen, 2015). Based on continued analysis of the original materials, concepts, and categories and their relations, four core categories are extracted: cost distance, initiative distance, control distance, and interactive distance. Finally, the online educational psychological distance model is constructed. That is, the relationship between core categories and categories, as shown in Table 5 and Figure 1.

Table 5. Correspondence between core categories and categories

			6 6
Number	Core category	Category	Core category interpretation
C1	Cost distance	B1 time	It shows that users perceive that online learning saves
		B2 space	time, space, and resources.
		B3 resources	
C2	Initiative distance	B4 cognition	It refers to the degree of users' cognition and expectations
		B5 expectations	with respect to online courses.
C3	Control distance	B6 self-control	It is manifested in users' perception of their own and
		B7 supervision	others' control in the learning process.
C4	Interaction distance	B8 emotions	It is the users' perception of emotion and interaction in
		B9 behavior	online learning.

The storylines for each psychological distance dimension are as follows:

C1 Cost distance: refers to the users' perception of the change in the cost of online education. Examples include eliminating teaching location restrictions, flexible course times, and availability of quality courses. This dimension can be regarded as the distance judgment of users with respect to time costs, space costs, and resource costs.

C2 Initiative distance: refers to the users' perception of their subjective initiative. Examples include the expected effect of online learning, i.e., the higher the expectation is, the more active the user is in learning, and vice versa. Users' cognition and judgment of their self-learning ability will also affect their learning initiative. This dimension can be regarded as the distance judgment regarding users' subjective initiative.

C3 Control distance: refers to the users' perception of control ability in the learning process. This control ability includes self-control and others' control, self-discipline in learning and supervision of the learning process by the teacher or the platform. This dimension can be regarded as the distance judgment regarding the users' control ability or as the management distance.

C4 Interactive distance: refers to the users' perception of interactivity in online learning. Such interactions include behavioral and emotional interactions. Behavioral interactions can motivate users to learn. Emotional interactions contribute to users' sense of trust in and dependence on the course, the teachers and the platform, fostering acceptance and recognition at the psychological level, improving learning stickiness and producing the ideal learning effect. This dimension can be viewed as the users' distance judgment regarding interactivity.



Figure 1. Core categories and categories

3.4. Saturation test

Theoretical saturation means that all the genera are fully developed in their attributes, dimensions, and forms of change. The saturation test is used to determine whether any new contributions can be made to the conceptualization through further data collection and analysis. If no new contributions emerge, the model is considered to have achieved theoretical saturation (Corbin & Strauss, 1990).

In this study, two-thirds of the data were randomly selected, and the remaining data were used to test the theoretical model of psychological distance in online education. Examples of theoretical saturation tests are as follows:

- a41: In the case of online teaching, it is not very convenient for teachers to communicate with students (interactive distance-behavior), and problems cannot be identified or solved in a timely manner via feedback (cost distance-time).
- a70: I think for me because I am, to an extent, forced to take offline classes (initiative distance-cognition), I don't like some courses, so I don't listen carefully to them. Generally, my learning state is not very good. In contrast, online learning is purposeful, and you will look for courses that you are satisfied with (cost distance resources). However, exam-oriented offline education is a little bit obsessive; it should be a little easier to stick to it (control distance supervision); online education is susceptible to other factors that require self-discipline (control distance self-discipline).
- a168: You can watch the playback online without worrying about keeping up with it, but it is common for there to be a lack of learning supervision (control distance self-control).
- a273: Sometimes the offline learning environment will be noisy, affecting the learning efficiency. On the internet, you can choose a quiet environment such as a library. In case of bad weather and for other reasons,

such as rain, online learning is relatively flexible and can be pursued in the same way in a dormitory (cost distance-space). There is also the problem of time allocation. When the study schedule is not flexible enough, online learning can be easily adjusted (cost distance-time).

• a372: Online teaching is separated by a screen, so there is no learning atmosphere in the classroom (interactive distance- behavior). Online, however, you can choose the courses of famous schools and teachers, and your knowledge absorption and learning effect can be guaranteed (initiative distance - expectation).

4. Discussion

In the context of competent network education, the psychological distance element model constructed in this paper summarizes four main categories: cost distance, initiative distance, control distance and interaction distance. The four dimensions are further explained and discussed below.

(a) Cost distance includes the time distance, space distance and resource distance. In the network education environment, learners attach importance to freedom in learning; i.e., they can learn anytime and anywhere. For example, "a9 I prefer online courses that can be watched anytime (time) and anywhere (space), so that I can make better use of my time." Therefore, the temporal and spatial distance here refers to the nonlimitation of time and space. The resource distance emphasizes the richness and optimization of resources. "a81 I usually take some basic courses online, but there are too many network resources for such courses, which leads to my confusion in choosing. If only the online education platform had intelligent recommendations according to personal needs (preferred)." Learners require the autonomy of online courses and want accuracy in course selection. Therefore, this study finds that the cost distance here is no longer simple space-time distance but the psychological perception of learners and their experience of the convenience of online courses.

(b) Initiative distance includes the cognitive distance and expectation distance. Learners' understanding and expectations of social needs, personal ability and curriculum will also affect their actual perception of the learning effect. For example, "a11, I found that professional knowledge in management is also very important, especially for engineering students like me (cognition). I expect to improve my work efficiency and ability (expectation) through online learning of management knowledge. Since there is no professional foundation, I prefer to choose online courses with richer cases and more classic theories (cognition). It can make me get more." Therefore, the active distance is also a subjective evaluation of learners of whether they can complete the curriculum task. It will directly affect the learning motivation, i.e., self-efficacy, of learners (Bandura, 1986).

(c) Control distance includes the self-control distance and supervision distance. It is found that while the learning autonomy of online courses is greater, it also requires a type of control constraint. This type of control includes both own active and third-party control to help learners complete the course smoothly." a113 In e-learning, I always procrastinate more. I must make a detailed learning plan to supervise myself (self-control) in order to successfully complete the course." "a54 I prefer to choose online courses with periodic assessment requirements (supervision), which is a good constraint for me." Therefore, the controlling distance can also be understood as a binding force on learners.

(d) Interaction distance includes emotional distance and behavioral distance. Learners hope to enjoy the flexibility and convenience of online course learning, and it is a reproduction of offline courses. There can be sufficient teacher-student interaction, student discussion, teacher attention and timely feedback on problems in the classroom. "a239 I enjoy having a heated discussion with CO learners on some issues in the learning forum, especially when my views or explanations are adopted by others (others' recognition of me), I will have a sense of achievement (self-identity)." The other is learners' recognition of the views, styles and cultures of teachers, learning teams and even learning platforms, which has gradually evolved into an emotional identity and preference, increasing the emotional distance. "a97 I like Professor Mengman's class (recognition of others). Taking her class, I will not only acquire knowledge but also feel more confident (self-identity)." Whether behavioral or emotional, such interaction also assists a learner in "identification with others" or "identification with me" to "self-identity." In turn, it improves his learning motivation and effect.

The uniqueness of this study is that it uses the grounded theory method, directly starts from actual observations with no previous assumptions, summarizes the experience from the original data, and subsequently establishes the theory. In the past, scholars used more quantitative analysis methods to conduct confirmatory research on educational psychological factors. Therefore, this study more completely shows the psychological perception

discovery of learners in e-learning (including cost distance, initiative distance, interaction distance and control distance) instead of only an analysis of the relationships among individual factors. In addition, this study finds that the four dimensions reflect the four core values needs of learners for online education. In other words, the learners' initial experience of online courses must have good convenience (cost distance) to give full play to the learning autonomy and accuracy of learners. However, the real learning motivation comes from the subjective evaluation of learners of their ability to complete online courses, i.e., self-efficacy (initiative distance). In addition, good user experience and user satisfaction also require learners to realize self-identity (interaction distance) in the links in online education. Moreover, the entire network teaching process requires a strong binding force (control distance) to help learners effectively complete learning tasks and ultimately achieve user satisfaction, as shown in Figure 2. Therefore, these four dimensions of psychological distance complement one another. Intelligent online education optimization can be considered from these four levels to improve the user experience and satisfaction of learners.





5. Conclusions and future work

• Four new dimensions of psychological distance in online education

Based on qualitative research employing the grounded theory approach, this study finds that psychological distance in online education includes four core categories, namely, four new dimensions. (i) Cost distance, or users' perception of distance in terms of time costs, space costs, and resource costs; (ii) Initiative distance, or users' subjective active distance perception; (iii) Control distance, or users' perception of their control ability in the process; (iv) Interactive distance, or distance perception based on interaction in online learning.

• In the context of network education, learners need to meet their four core values from the level of psychological perception.

Cost distance is learners' specific perception of the convenience of online courses, which is expressed as the primary level of online education user experience satisfaction. Initiative distance is the perception of self-efficacy, that is, the level of learners generated learning motivation through cognition and expectation. Interaction distance is the process of achieving self-identity through behavioral and emotional interaction, which is expressed as the advanced level of online education user experience satisfaction. Control distance is a self or third-party binding perception to promote learners to smooth completion of the course. Therefore, these four dimensions of psychological distance interact with each other and ultimately form user experience and satisfaction in online education.

As exploratory research, this paper uses grounded theory to obtain four dimensions of learners' psychological distance against the background of network education to optimize user experience and satisfaction in intelligent network education. Combined with the research of this paper, the following problems merit further discussion.

- This paper categorizes the psychological distance of online education learners into four dimensions. With the development of online education, whether there are other dimensions remains to be further discussed. In addition, the conclusion of this study still lacks empirical testing.
- Some aspects of the application of grounded theory in this study need to be further improved in the specific research process. For example, how should the category be defined in the formation process from concept to category, and what is the theoretical basis of the defined category? I believe that with the continuous improvement of grounded theory, its research will become more standardized.

Acknowledgement

This study is supported by the Development Program for Young and Middle-aged Key Teachers of Chengdu University of Technology 2020.

References

Amit, E., Algom, D., & Trope, Y. (2009). Distance-dependent processing of pictures and words. *Journal of Experimental Psychology: General*, 138(3), 400. https://doi.org/10.1037/a0015835

Afrouz, R., & Crisp, B. R. (2021). Online education in social work, effectiveness, benefits, and challenges: A Scoping review. *Australian Social Work*, 74(1), 55-67.

Bandura, A. (1986). Social foundations of thought and action: A Social cognitive theory. Prentice-Hall, Inc.

Bar-Anan, Y., Liberman, N., & Trope, Y. (2006). The Association between psychological distance and construal level: Evidence from an implicit association test. *Journal of Experimental Psychology: General*, *135*(4), 609. https://doi.org/10.1037/0096-3445.135.4.609

Bar-Anan, Y., Liberman, N., Trope, Y., & Algom, D. (2007). Automatic processing of psychological distance: Evidence from a Stroop task. *Journal of Experimental Psychology: General*, *136*(4), 610–622. https://doi.org/10.1037/0096-3445.136.4.610

Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., & Berta, R. (2013). Assessment in and of serious games: An Overview. *Advances in Human-Computer Interaction*, 2013. https://doi.org/10.1155/2013/136864

Boroditsky, L., & Ramscar, M. (2002). The Roles of body and mind in abstract thought. *Psychological science*, 13(2), 185-189.

Breves, P., & Schramm, H. (2021). Bridging psychological distance: The Impact of immersive media on distant and proximal environmental issues. *Computers in Human Behavior*, *115*, 106606. https://doi.org/10.1016/j.chb.2020.106606

Bullough, E. (1912). Psychical distance as a factor in art and an aesthetic principle. *British journal of Psychology*, 5(2), 87-118.

Cantabella, M., R Martínez-Espaa, B López, & Muoz, A. (2020). A Fine-grained model to assess learner-content and methodology satisfaction in distance education. *International Journal of Interactive Multimedia and Artificial Intelligence*, 6(4), 87-96. https://doi.org/10.9781/ijimai.2020.09.002

Chen, H., & Guibing, H. E. (2014). The Effect of psychological distance on intertemporal choice and risky choice. Acta Psychologica Sinica, 46(5), 677–690.

Chen, H., & Li, S. (2018). Measuring the psychological distance between an organization and its members—The Construction and validation of a new scale. *Frontiers in Psychology*, *8*, 2296. https://doi.org/10.3389/fpsyg.2017.02296

Corbin, J., & Strauss, A. (1990). Grounded theory method: Procedures, canons, and evaluative procedures. *Qualitative Sociology*, 13, 13-21.

Corbin, J., & Strauss, A. (2014). Basics of qualitative research: Techniques and procedures for developing grounded theory. Sage publications.

Creswell, J. W. (2006). *Qualitative inquiry and research design: Choosing among five approaches* [with CD-ROM] (2nd ed.). Sage publications.

Dhar, R., & Kim, E. Y. (2007). Seeing the forest or the trees: Implications of construal level theory for consumer choice. *Journal of Consumer Psychology*, 17(2), 96-100.

Engebretson, D. E. (1973). Human territorial behavior: The Role of interaction distance in therapeutic interventions. *American Journal of Orthopsychiatry*, 43(1), 108-116. https://doi.org/10.1111/j.1939-0025.1973.tb00790.x

Farrell, B., Ward, N., Jennings, B., Jones, C., Jorgenson, D., Gubbels-Smith, A., Dolovich, L. & Kennie, N. (2016). Participation in online continuing education. *International Journal of Pharmacy Practice*, 24(1), 60-71.

Fiedler, K. (2007). Construal level theory as an integrative framework for behavioral decision-making research and consumer psychology. *Journal of Consumer Psychology*, *17*(2), 101-106.

Foley, G., & Timonen, V. (2015). Using grounded theory method to capture and analyze health care experiences. *Health* Services Research, 50(4), 1195-1210.

Gasson, S., & Waters, J. (2013). Using a grounded theory approach to study online collaboration behaviors. *European Journal of Information Systems*, 22(1), 95-118.

Green, C. (2013). Relative distancing: A Grounded theory of how learners negotiate the interprofessional. *Journal of Interprofessional Care*, 27(1), 34-42.

Glaser, B. G., & Strauss, A. L. (1967). The Discovery of grounded theory: Strategies for qualitative research. Aldine Publishing Company.

Ho, C. K. Y., Ke, W., & Liu, H. (2015). Choice decision of e-learning system: Implications from construal level theory. *Information & Management*, 52(2), 160-169. https://doi.org/10.1016/j.im.2014.07.003

Holt, N. L., Neely, K. C., Slater, L. G., Camiré, M., Côté, J., Fraser-Thomas, J., MacDonald, D., Strachan, L., & Tamminen, K. A. (2017). A Grounded theory of positive youth development through sport based on results from a qualitative meta-study. *International review of sport and exercise psychology*, *10*(1), 1-49.

Horvath, P. (2018). The relationship of psychological construals with well-being. New ideas in psychology, 51, 15-20.

Kožuh, I., Jeremić, Z., Sarjaš, A., Bele, J. L., Devedžić, V., & Debevc, M. (2015). Social presence and interaction in learning environments: The effect on student success. *Journal of Educational Technology & Society*, 18(1), 223-236.

Krippendorff, K. (2018). Content analysis: An Introduction to its methodology. Sage publications.

Kwon, S. Y. (2011). The Investigation of the relationships among social presence, learners' characteristics, perceived effects of discussion, perceived learning and satisfaction in online learning. *The Journal of Educational Studies*, 42(3), 55-82.

Lee, H. J., Hong, Y., & Choi, H. (2017). Perceptions of tutoring roles and psychological distance among instructors, tutors and students at a Korean university. *Higher Education Research & Development*, *36*(1), 143-157.

Lee, J. E. (2010). Psychological distance between students and professors in asynchronous online learning, and its relationship to student achievement & preference for online courses. *Educational Technology International*, 11(2), 123-148.

Li, S., Chen, H., Huang, X., Hou, C., & Chen, F. (2019). Chinese public response to occupational safety and health problems—A study based on psychological distance. *International journal of environmental research and public health*, *16*(11), 1944. https://doi.org/10.3390/ijerph16111944

Liberman, N., Trope, Y., & Stephan, E. (2007). Psychological distance. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (pp. 353–381). The Guilford Press.

Liberman, N., & Trope, Y. (1998). The Role of feasibility and desirability considerations in near and distant future decisions: A test of temporal construal theory. *Journal of personality and social psychology*, 75(1), 5-18. https://doi.org/10.1037/0022-3514.75.1.5

Liu, Q., Zhang, X., Huang, S., Zhang, L., & Zhao, Y. (2020). Exploring consumers' buying behavior in a large online promotion activity: The role of psychological distance and involvement. *Journal of theoretical and applied electronic commerce research*, *15*(1), 66-80.

Lofland, J. (2006). Analyzing social settings: A Guide to qualitative observation and analysis (4th revision ed., Vol. xxi). Wadsworth.

Maglio, S. J., & Polman, E. (2014). Spatial orientation shrinks and expands psychological distance. *Psychological Science*, 25(7), 1345-1352.

Matcha, W., Gašević, D., & Pardo, A. (2019). A Systematic review of empirical studies on learning analytics dashboards: A Self-regulated learning perspective. *IEEE Transactions on Learning Technologies*, *13*(2), 226-245.

Neroni, J., Gijselaers, H. J., Kirschner, P. A., & de Groot, R. H. (2015). The Adult Learning Open University Determinants (ALOUD) study: Biological and psychological factors associated with learning performance in adult distance education. *British Journal of Educational Technology*, *46*(5), 953-960. https://doi.org/10.1111/bjet.12288

Orlando, J. (2015). Gameful design in online education. Online Classroom, 15(11), 5-7.

Patton, M. Q. (2014). Qualitative research & evaluation methods: Integrating theory and practice. Sage publications.

Sahoo, K. C., Hulland, K. R., Caruso, B. A., Swain, R., Freeman, M. C., Panigrahi, P., & Dreibelbis, R. (2015). Sanitation-related psychosocial stress: A Grounded theory study of women across the life-course in Odisha, India. *Social science & medicine*, *139*, 80-89.

Sbaraini, A., Carter, S. M., Evans, R. W., & Blinkhorn, A. (2011). How to do a grounded theory study: A Worked example of a study of dental practices. *BMC Medical Research Methodology*, 11(1), 1-10.

Sungur, H., Van Koningsbruggen, G. M., & Hartmann, T. (2017). Psychological distance cues in online messages. *Journal of Media Psychology*, *31*(2). https://doi.org/10.1027/1864-1105/a000229

Tolman, E. C. (1932). Purposive behavior in animals and men. University of California Press.

Trope, Y., & Liberman, N. (2003). Temporal construal. Psychological review, 110(3), 403-421.

Vayre, E., & Vonthron, A.-M. (2017). Psychological engagement of students in distance and online learning: Effects of self-efficacy and psychosocial processes. *Journal of Educational Computing Research*, 55(2), 197-218.

Villagrá-Arnedo, C. J., Gallego-Durán, F. J., Llorens Largo, F., Satorre Cuerda, R., Compañ, P., & Molina-Carmona, R. (2020). Time-dependent performance prediction system for early insight in learning trends. *International Journal of Interactive Multimedia and Artificial Intelligence*, 6(2) 112-124. https://doi.org/10.9781/ijimai.2020.05.006

Vaughn, L. M., & Baker, R. C. (2004). Psychological size and distance: Emphasising the interpersonal relationship as a pathway to optimal teaching and learning conditions. *Medical education*, 38(10), 1053-1060.

Warren, J. A., & Nash, A. (2019). Using expressive arts in online education to identify feelings. *Journal of Creativity in Mental Health*, 14(1), 94-104.

Weidlich, J., & Bastiaens, T. J. (2018). Technology matters-The impact of transactional distance on satisfaction in online distance learning. *International Review of Research in Open and Distributed Learning*, 19(3). https://doi.org/10.19173/irrodl.v19i3.3417

Zhbanova, K. S., & Rule, A. C. (2014). Construal level theory applied to sixth graders' creativity in craft constructions with integrated proximal or distal academic content. *Thinking Skills & Creativity*, *13*, 141-152.
Design and Development of a Blockchain-Based Secure Scoring Mechanism for Online Learning

Cheng-Ting Tsai¹, Ja-Ling Wu², Yu-Tzu Lin^{3*} and Martin K.-C. Yeh⁴

¹Department of Electrical Engineering, National Taiwan University, Taiwan // ²Department of Computer Science, National Taiwan University, Taiwan // ³Graduate Institute of Information and Computer Education / Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, Taiwan // ⁴Information Sciences and Technology, Pennsylvania State University Brandywine, USA // chtitsai@gmail.com

// wjl@cmlab.csie.ntu.edu.tw // linyt@ntnu.edu.tw // kqy1@psu.edu

*Corresponding author

ABSTRACT: With the rapid increase of online learning and online degree programs, the need for a secure and fair scoring mechanisms in online learning becomes urgent. In this research, a secure scoring mechanism was designed and developed based on blockchain technology to build transparent and fair interactions among students and teachers. The proposed scoring mechanism was implemented by employing the Ethereum blockchain and its three autonomous smart contracts. The robustness and feasibility of the system was then verified with experiments. The resulting system is shown to be superior to existing online learning systems because it prevents answer tampering. In addition, fairness can be improved with blockchain protocols and a collaborative scoring policy. Lastly, this system helps manage interactions among students and teachers during the process of educational assessment, and encourages all on-chain members to trust the online learning process. These advantages improve peer evaluation and self-directed learning that are essential for a student-centered and collaborative learning environment.

Keywords: Blockchain, Ethereum, Cryptography, Online learning, Online assessment

1. Introduction

Over the last few years, researchers and application developers took blockchain technology more seriously since its most well-known realization, Bitcoin, was introduced by Satoshi Nakamoto in 2008 (Nakamoto, 2008). The success of Bitcoin shows that blockchain techniques do contribute to the stability and liveliness of a system where data and executive activities are decentralized—supervised and maintained by all members of the chain. This decentralization feature makes each interaction immutable, secure, and transparent. This explains why blockchain technology has been applied to various fields, such as profit sharing and credit scoring (Jain et al., 2019), where members are treated equally, and when the legitimacy of transferred information has to be considered seriously. In the field of education, security issues for teacher-student interaction were often not discussed in a traditional learning environment. However, as online learning and online degree programs are growing rapidly (Porter, 2015), a reliable and secure scoring mechanism for learning management systems is required to prevent possible cheating and guarantee fair and accurate assessment results. In fact, COVID-19 has accelerated teaching mode from physical to online format (Pavlov & Katsamakas, 2021), the need of secure and private learning systems becomes urgent.

Traditional online learning platforms might suffer from security vulnerability due to the lack of security mechanisms and unequal privileges. Existing research focused more on preventing cheating during online examinations by applying biometrics technologies (Apampa et al., 2010; Traoré et al., 2017; Sabbah, 2017) or multi-factor authentication (Urosevic, 2019) to increase the security during examination, introducing a liveremote human proctor for exam monitoring (Lilley et al., 2016), or proposing a conceptual framework to provide guidelines for online examinations (Ngqondi et al., 2021). However, answer tampering after tests or subjective scoring biases have not been addressed. Apparently, if the on-platform activities are not traceable, the system may not be trustworthy in ensuring fair and unforged teacher-student interactions. Another common challenge in assignment/examination scoring is scoring open-ended questions (e.g., essay questions and calculation problems). Teachers might have different opinions and biases, which leads to disagreement about the assessment results. There is still limited research studying how to develop security models for online learning. Although previous research has tried to develop an architecture of trustworthy web services for secure assessment for collaborative learning (Caballé et al., 2017), such architecture was built for grid infrastructure. To solve the problem of scoring biases, collaborative scoring is a possible solution because it can include various opinions from different scorers, which is common in collaborative and project-based learning. It, however, might cause the bandwagon effect if there is no proper scoring mechanism. Therefore, it is required to design a secure scoring mechanism for fair and effective scoring. Moreover, with the increasing number of online courses, more and

more educational records will be stored and shared virtually over an array of networks. Invariably, we are facing the risks associated with hackers and other unethical actors. Blockchain technology can help secure and protect data in this new education model for its ability to combine information security and share data virtually to conduct learning among a wide range of networks.

In this paper, a blockchain-based assignment scoring mechanism is implemented to achieve fair and transparent teacher-student interactions during assessment, with which on-chain members are anonymous and their interactive activities are immutably traceable. To demonstrate its feasibility and applicability, we implemented the system on the Ethereum architecture along with multiple cryptography algorithms. Our teacher-student interaction model was designed to make all members equal and remove flaws in the scoring system, such as biases (by teachers or teaching assistants) and answer tampering (by students). Teachers can only uncover the students' identities at the end of the course to ensure the fairness of scoring. Further, three autonomous smart contracts were designed to guarantee the fairness and the efficiency of assessment. Finally, the proposed mechanism was implemented and the feasibility and robustness were examined by experiments.

2. Literature review

2.1. Online Learning Management System

An online Learning Management System (LMS) is a platform providing services of administration, assessment, reporting, automation, and delivery of educational courses, on which interactions among teachers and students might affect the effectiveness of online learning (Wright, 2014). With the rapid growth of online education, assessment for online assignments/examinations becomes a significant issue. However, the architecture of the LMS might not be secure enough to prevent misconducts, and fairness of assessment might be affected by student and teacher perceptions. Previous research has investigated possible solutions to increase security and fairness of online learning. Some studied how to strengthen authentication and identification systems to increase examination security by employing biometrics technologies (Apampa et al., 2010; Traoré et al., 2017; Sabbah, 2017) or multi-factor authentication (Urosevic, 2019). Examination monitoring is a solution to ensure fairness, which can be implemented by a live-remote human proctor (Lilley et al., 2016). Instead of only considering cheating prevention, some systems focus on improving the security of the LMS architecture, including an architecture of secure assessment by trustworthy web services (Caballé et al., 2017), a conceptual framework to provide guidelines for online examinations (Ngqondi et al., 2021), and a secure assessment management system based on cryptography protocols (Castella-Roca et al., 2006). However, the methods are either too complex to implement in a general LMS or only suitable for specific infrastructure. In addition, some illegal behaviours, such as answer tampering after the examination, are ignored in those studies.

2.2. Blockchain in education

Blockchain can be used to carry and transfer any valuable assets, such as currency, copyrighted materials, knowledge, and records. In education, there are many valuable information, including research data, experimental records, scores, credits and certificates of degrees whose management, security and fairness are necessary and extremely important for all stakeholders. Therefore, blockchain might be a suitable vehicle to bring benefits to educations (Chen et al., 2018; Skiba, 2017; Hernandez-de-Menendez et al., 2020) and makes management of all the students' and educators' information fairly and efficiently. For online education platforms, such as MOOCs, where students and educators come from different places of the world to achieve their own educational goals. Then the learning environment becomes more diverse, establishing trust between each member becomes a significant and challenging task.

The report of the European Commission's Joint Research Centre (JRC) suggests that issuing certificates is an important application for education (Grech & Camilleri, 2017), which involves tracking learning data (e.g., portfolio and achievements) to approve certificates (Raimundo & Rosário, 2021). Many studies focused on using blockchain to manage, share, and verify degrees/certificates and credits (Sharples & Domingue, 2016; Turkanović et al., 2018), or research results and data (Hoy, 2017). Some research studied secure assessment mechanisms for online learning (Lam & Dongol, 2020; Sudaryono et al., 2020). However, most of the works paid attention to managing the "post stage" of educational activities such as recording and sharing certificates, diplomas, and grades between institutions to protect the "results" of learning and assessment processes. As illustrated in Table 1, although recent research has proposed frameworks and algorithms for secure certificate verification or grade management, there is still limited research exploring effective algorithms for security and

privacy "during" learning, not to mention the consideration of pedagogical features (e.g., collaborative learning or scoring bias). Additionally, implementation of the system and its performance evaluation are still lacking.

<i>Table 1</i> . Research of blockchain in education						
Educational applications	Research	Features				
Research results and data management	Document management (Das et al., 2021)	Use smart contracts to track, manage, and store documents to facilitate approval flows and apply public-key cryptography to facilitate data confidentiality and integrity				
Certificate validation and management	Certification for e-learning (Li et al., 2019) Higher education credit management (Turkanović et al., 2018)	Store e-learning data in a Merkle tree and manage credits using a public blockchain Use the DPoS consensus protocol to achieve globally unified viewpoint for students and higher education institutions				
	Storing and managing degree information (Nazare et al., 2016)	Store certificate data in a Merkle tree while preserving the ability for individual users to access their own certificates				
Assessment	Automate assessment for e-learning (Lam & Dongol, 2020)	Use smart contract to send test files for automated marking and grade calculation and storage				
	Grade management (Sudaryono et al., 2020)	Prevent modification of grades by recording all processes in the blockchain				
	Grades storage and calculation for e- learning (Li et al., 2019)	Manage grades by allocating e-learning voucher to ensure the credibility based on a private blockchain				

2.3. Ethereum

Ethereum is one of the blockchain architectures introduced between 2013 and 2014, devoting to establish a global and most completed blockchain system. Ethereum is very popular and considered to be a huge breakthrough in blockchain technology. One of the important contributions of Ethereum is its introduction of Smart Contract, a computerized transaction protocol that executes the terms of a contract and is written by a specific programming language, such as Solidity (Dannen, 2017). Smart Contract can be independently and autonomously executed by nodes on an Ethereum network using virtual machines, which are called Ethereum Virtual Machines (EVMs). The Turing-completeness of Smart Contract allows Ethereum blockchain to be applied to many complex tasks, such as funding, supply chaining, bidding, and even signing another contract. These features transformed blockchain technology from a purely distributed system that can only send transactions (Jansen et al., 2019) to a completed decentralized architecture that can perform complex tasks and transfer virtual currencies. Ethereum is also open sourced so that everyone can join and research on it, or build his or her own designed private Ethereum-based chains. Therefore, if one wants to design a blockchain system to fulfill some complex use cases using Smart Contract, Ethereum is one of the best platforms. Considering its features, our work is realized based on the Ethereum architecture.

3. Scope and assumptions

3.1. Regulations and account management

This study aims at developing the required system for a practical educational scenario. Therefore, the proposed assignment scoring mechanism is expected to operate properly under the supervision of an educational institution or an online learning system, where regulations are made to restrict both students and teachers from sabotaging the system. This may seem to centralize the system; however, the operations of the system are designed not to be interfered by the administrator. This means the system is decentralized running by the students and teachers who follow the regulations under the administrator's supervision. Furthermore, the administrator has to verify the status of students and teachers after they signed in the system and intervenes between students and teachers only when some disputes against the preset rules occurred.

In each quarter or semester, every qualified member, such as teacher, student, and teaching assistance, will respectively receive an address that points to the corresponding account used in the assignment/scoring system, from the administrator. After registration, the administrator gives teachers their student lists that contain student accounts associated with the corresponding classes (to prevent non-registered students from joining the courses without permission) and the students' IDs to identify that the students did take the classes at the end of the quarter/semester (see Section 4.4 for details). Note that the correspondence between accounts and students' ID remains in secret (see Figure 1). That is, the teacher will never know which student owns a specific account until the course is finished. The administrator uses the accounts to track and supervise members' behaviors to enhance the stability and liveliness of the system. Offenders are suspended or punished according to the regulations or even laws depending on the severity of violation.

To make the system highly reliable and functional, supervisions and regulations are necessary. However, the system will still operate in a decentralized manner due to the nature of blockchain. Once the system starts, it will be maintained and verified by every on-chain member and its operation will be almost impossible to interfere with or temper the data stored on it, not even by the administrator.



·······

3.2. Computational power

In the purposed work, the computational power is assumed to be uniformly distributed among all involved members. That is, each member joins the consensus mechanism and has an equal chance and responsibility to create a new block and maintain the liveness of the system. Even though some members do have better computational power than others, it is assumed that no one member will gather enough computational power to conspire against or even sabotage the system.

4. Blockchain-based scoring mechanism

4.1. Basic member interaction models

In every scenario of education, an interaction between students and teachers is a must. Our simplest model aims at simulating the interaction between students and teachers through a blockchain architecture. To do so, the following three functional modules must be defined: assignment delivery and submission, scoring results delivery, and class information announcement. A teacher can deliver assignments or announce information to students by simply sending transactions with messages. Following the same principle, students can submit their assignment answers. However, it does not make sense to put all messages (e.g., answers to assignments) directly on a transaction because a blockchain is a transparent system, which means every on-chain member can see the content of any validated transaction. In short, submitting assignment answers in its plaintext form would result in exposing students' answers to everyone. Therefore, messages that are not suitable to be publicized must be encrypted before sending. There are various ways of encrypting messages to transmitted securely. In this paper, the Rivest–Shamir–Adleman (RSA) algorithm (Calderbank, 2007; Rivest et al., 1978), one of the most widely used encryption methods that is easy to implement and very hard to be cracked, is applied. To use this encryption algorithm (Figure 2), the teacher needs to generate a key pair (a public key and a private key) and sends the public key to the students along with the assignment.

Students then use the RSA algorithm with the shared public key to encrypt their answers and send the ciphertexts to the teacher. The messages include answers for the assignment and student's identity (Section 4.4) so that ciphertexts look different even if the answers are the same. Such design can be effective to prevent plagiarism.

Finally, the teacher can restore the students' assignment answers by decrypting the ciphertexts with the private key.



By this approach, our system not only keeps the messages in secret but also prevents answer tamping. All students and teachers put their trust in this model, every system player has fair rights and legal duties to maintain and interact with the model. All system activities are easy to be tracked and supervised; therefore, the deletion or modification of any content of the announcement, assignment, and assignment submissions is nearly impossible. Consequently, the proposed model can build a secure and fair online course platform by using blockchain.

4.2. The role of teaching assistants

Teaching assistants (TAs) are often recruited to help run large courses and distribute assignments. One of the most common tasks that a TA is required to do is to grade assignments and tests. Thus, our model is extended to take the interactions among students, TAs, and teachers into account. As shown in Figure 3, the bottom half of the new model is similar to Figure 2. The only difference is that grading submitted tasks is now done by TAs. In this case, teacher still needs to deliver assignments to students and assign a TA to each student and gives a proof to evidence that the assignment is indeed released by the teacher. After receiving the verified assignment, TAs and students interact with each other accordingly.





4.3. Collaborative scoring

In real online learning, teachers sometimes design open-ended questions (e.g., essays) to assess students' understanding. It is challenging to assess this type of questions. On the one hand grading using one scorer could possibly cause a bias. On the other hand, if more scorers are included, there might be disagreements among scorers with various perspectives. Therefore, in many high-stake examinations, more than one scorer is involved in scoring to avoid biases. In this study, we also design a secure scoring mechanism for collaborative scoring.

In our collaborative scoring system (shown in Figure 4), students, the responsible teacher, and multiple scorers are invited in the assessment process. The assignments still need to be given by the responsible teacher to prove its legality. Students have to submit their answers in ciphertexts to all scorers. After the deadline, the teacher shares the assignments, students' answers, and sometimes the teacher's remarks (this is optional), to other scorers, while scorers use the ciphertexts and the public key received from the teacher to verify the plaintexts to ensure the plaintexts have not been tempered. Finally, scorers will send their scores to the teacher and corresponding students, so that the final scores can be calculated based on a preset weighting. The teacher sends two scores to each student, one is the teacher's score as a judgement and the other is the final (collected and weighted) score sent to be recorded and verified. Once again, because messages are trackable in blockchain, forging scores become very difficult.





Another issue is regarding the timing of receiving students' submissions by the scorers. To prevent the scorers from discussing submissions with others, which may introduce the scoring bias, the assignments should be kept secure during the submission stage. However, data on blockchain are transparent to all on-chain members. Therefore, an RSA public-private key-pair can be employed to make the assignments secure from scorers at the submission stage. But the process might be inefficient: all students generate their own key-pairs and the teacher encrypts the assignment for each student with the student's key individually. To improve the usability and simplify the process for both teacher and students, we chose to have a system administrator generate the key-pairs and distribute the public keys to the teacher and the private keys to the students. The teacher then can send encrypted assignments and only the corresponding students can decrypt. When the scoring stage starts, the scorers receive the plaintext of the assignment and students' answers for scoring. The scorers then can use the public key to verify and ensure that the answers are not tempered.

By this approach, the scores for opened questions will be more reliable. Additionally, the scorers in this model are anonymous, so that each scorer can judge the quality of answers without being affected by others (e.g., the owner of the answers). As a result, the proposed approach can improve quality and fairness. Specifically, for extremely high-stake assessments such as examinations that are directly affecting the issuing of certificates or the college entrance qualifications, the abovementioned method is believed to be a trustworthy way for establishing credibility in scoring.

4.4. Authentication

On a blockchain, every member is identified by a hash string and their activities are thus anonymized. In other words, this property allows students to take courses without giving up their identity. It also ensures that teachers treating their students equally. However, the teacher needs to recover students' IDs to give final scores. To achieve this, a Shamir's Secret Sharing algorithm (Shamir, 1979) and Chaotic Cryptography algorithm (Kocarev & Lian, 2011) based authentication scheme are used and are discussed in the rest of this section.

In the beginning of the course, by using Chaotic Cryptography, each student generates his/her secret codes by encrypting his/her student ID with the chosen password and segments his/her secret codes into secret pieces by using the Shamir's Secret Sharing algorithm (Figure 5). Students then send each one of their secret pieces together with their submitted assignments to the teacher so that the teacher can eventually find their secret codes out (Figure 6). The *t*-out-of-N, (N,t)-Threshold Shamir's Secret Sharing algorithm is adopted. This algorithm initially segments the secret into N pieces, and the secret can later be recovered if at least t out of the N pieces are retrieved. In our system, the parameter N is set to the total number of assignments in a course, and t is the least number of assignments that a student has to submit, which is determined by the teacher.



Figure 6. Teacher recovers a student's secret code by retrieving secret pieces sent together with the submitted assignment



Figure 7. Teacher identifies a student with the password (R_1) and the associated secret code



In considering the fact that it is not realistic to expect every student to submit every assignment on time due to unpredicted reasons such as missing the deadline or cannot complete the assignment, the teacher should still be able to recover the secret codes with only a part of secret pieces. In some cases, students are not asked to turn in all assignments but at least a minimum number of assignments. If a student fails to fulfill the minimum requirement, the teacher won't be able to recover the secret codes to identify the student, and thus no final score will be given to the student. In contrast, if a student can prove the efforts that they put to the course, the final score should still be given even some of the submissions are missing.

Integrating the secret sharing algorithm with the system makes it closer to the needs of real application scenarios. However, the adopted (N,t)-Threshold scheme also imply that the teacher can obtain students' identity before the course is completed. To solve this problem, Chaotic Cryptography is applied to protect students' privacy. A teacher can never find out the student ID within the secret codes without knowing the password set by the student. That is, a student's identity will remain in secret before the student sent out the final key information, i.e., the password, to the teacher, at the end of the course (Figure 7). Additionally, because there is no student ID shared on the blockchain, a student's identity behind a given account is safe and remains unknown to the other members. By combining the above schemes, a teacher can identify the students enrolled in the course and set

some rules (such as the values of N and t) for the course while the students' privacy is well protected by their privately set passwords.

5. Smart contracts

Smart contract is a crucial feature of blockchain that uses designed protocol to autonomously run on decentralized networked nodes for achieving various complex tasks. Once deployed, a smart contract acts as a fair and transparent arbiter to deal with every request from its users. In education, there are many complex situations that a smart contract can be applied to make things easier. For example, it can be used to collect group lists or act as a billboard to announce information. It is worth mentioning that a well-designed smart contract can also replace TAs for completing tasks that follow unambiguous rules, such as scoring assignments. Smart contracts guarantee tasks can be done objectively comparing with TAs who may have specific personal opinions on certain students. To ensure the fairness and transparency of scoring, three approaches to score assignments or examinations by a smart contract are proposed: *peer evaluation, automatic scoring by a smart contract*, and *collaborative scoring*.

As mentioned in Section 4.1, sending answers in plaintext equals sharing answers to everyone on the blockchain, which is certainly not ideal. However, using a smart contract to decrypt a ciphertext is very difficult and costly due to the complexity of crypto algorithms. What is worse is uploading a private key to the blockchain not only has to pay the cost for storing large random numbers but also reveal the private key to all on-chain members. Practically, it is not trivial to avoid mistakes when embedding a huge-size message into a transaction. Therefore, a better way to protect information security is to use the smart contract to directly verify the ciphertext with the aid of various commitment schemes instead of decrypting it back to plaintext and then score. As illustrated in Figure 8, when an assignment is announced, students need to upload their answers in a ciphertext form before the deadline. The secure hash algorithm used to obtain the ciphertext should also be supported by the smart contract to truly optimize the efficiency. In our work, the keccak256 hash algorithm, which is a callable function to Solidity language, is adopted.



Figure 8. Schematic diagram of smart contract based scoring by using mutual exchanging mechanism

Similar to the basic model, the plaintext should contain an extra message, which is denoted as R in Figure 8, to prove the student's identity at the end of the course and prevent answer-tamping or assignment-copying flaws to ensure every student will get a unique hash value even if their answers are the same. After the deadline, students upload their answers together with message R. The integrity of answers can be proved by the smart contract via checking if the hash of the plaintext matches the uploaded ciphertext. Once it is confirmed the answers in plaintext can be scored manually or automatically by the smart contract.

5.1. A smart contract to support peer evaluation

Peer evaluation is to let students score other student's assignments. On a blockchain, no student is able to know the owner of other addresses. That is, a student does not know whose assignment he or she is grading, and therefore, will reduce the chance of cheating. In addition, the smart contract is designed to make the assignment of peer evaluation randomly. Even if a student shared his or her address with friends, there is no guarantee that they will be paired, especially when there are many students enrolled in the course.

The proposed smart contract requires four basic functions for peer evaluation to work: (1) submit ciphertext, (2) start scoring, (3) submit plaintext, and (4) fetch the assignment that needs to be corrected/scored. As shown in Table 2, initially, only function (1) is activated for students to submit their ciphertext (with commitment) while functions (3) and (4) remain disabled until the teacher calls function (2) and uploads the solutions or the rubrics after the assignment deadline. At this point, function (1) is also disabled to prevent students from submitting new answers. The algorithm of scoring by exchanging is illustrated in Figure 9.

1000000000000000000000000000000000000

Functions	(1) Submit ciphertext	(2) Start scoring	(3) Submit plaintext	(4) Fetch assignment
After (2) is called	Disable	Disable	Enable	Enable

1:	// Teacher creates Contract	
2:	procedure CONSTRUCTOR()	30: // Students upload plaintexts
3:	$owner \leftarrow sender$	31: procedure SETPLAINTEXT(plaintext)
4:	$flag \leftarrow false$	32: if sender exist and flag then
5:		33: $map(sender).plantext \leftarrow plaintext$
6:	// Students upload ciphertexts before deadline	
7:	procedure SETSTUDENTS(ciphertext)	34:
8:	if <i>flag</i> then return	35: // Students receive assignment to correct
9:		36: procedure GETASSIGNMENTTOCORRECT()
10:	// Update or Add a student into Map	37: if sender exist and flag then
11:	if <i>sender</i> already exist then	38: $result \leftarrow map(sender).correct for$
12:	$map(sender).hw \leftarrow ciphertext$	39: $a \leftarrow map(result).addr$
13:	else	40: $c \leftarrow map(result).hw$
14:	$newStudent \leftarrow map(sender)$	41: $p \leftarrow map(result).plaintext$
15:	$newStudent.addr \leftarrow sender$	42: return (a, c, p)
16:	$newStudent.hw \leftarrow ciphertext$	
17:	addresses.push(sender)	
18:		
19:	// Teacher exchange assignment before deadline	
20:	procedure SETSTARTCORRECT()	
21:	if sender is not owner or flag then return	
22:		

5.2. A smart contract to support automatic scoring

 $map(addr).correctfor \leftarrow addresses[0]$

 $map(addr).correct for \leftarrow addr + 1$

for each *addr* in *addresses* do

 $\mathbf{if} \ \mathbf{last} \ addr \ \mathbf{then}$

23:

24: 25:

26:

27:

28: 29: $flag \leftarrow true$

else

Another way to make scoring fairly to every student is to let the smart contract grades the submissions, as illustrated in Figure 10. This kind of smart contracts also have four basic functions: (1) submit ciphertext, (2) start scoring, (3) submit plaintext, and (4) fetch scoring results. When calling function (2), the teacher receives the actual answers and changes the status of the other three functions as shown in Table 2. The smart contract then scores the answer once the student calls function (3) and uploads the plaintext that matches the verified ciphertext uploaded by function (1). Finally, function (4) allows all enrolled students to see the results of their assignments.

Figure 10. Schematic diagram of smart contract based automatic scoring system



Comparing with the smart contract proposed in Section 5.1, this approach simplifies students' work loads and guarantees fairness to all students because the smart contract autonomously grades every submission. However, to make this method feasible, both the solutions and their forms in the plaintext domain must be fixed to make sure that the smart contract can match or extract correct solutions from the plaintexts. For this reason, the teacher must upload answer keys rather than guidelines or rubrics. Thus, scoring essay is hard to achieve using this approach. The automatic scoring algorithm is illustrated in Figure 11.

Figure 11. Algorithm: Automatic scoring

1:	// Teacher creates Contract
2:	procedure CONSTRUCTOR()
3:	$owner \leftarrow sender$
4:	$flag \leftarrow false$
5:	
6:	// Students upload ciphertexts before deadline
7:	procedure SETSTUDENTS(ciphertext)
8:	if <i>flag</i> then return
9:	
10:	if sender already exist then
11:	$map(sender).hw \leftarrow ciphertext$
12:	else
13:	$newStudent \leftarrow map(sender)$
14:	$newStudent.addr \leftarrow sender$
15:	$newStudent.hw \leftarrow ciphertext$
16:	addresses.push(sender)
17:	
18:	// Teacher starts to correct assignment
19:	procedure SETSTARTCORRECT(answer)
20:	if <i>sender</i> is not <i>owner</i> or <i>flag</i> then return
21:	
22:	// Extract each answer
23:	$flag \leftarrow true$
24:	for each delimiter in <i>answers</i> +1 do
25:	ans.push(answers.split(delimiter))

26:	
27:	// Students upload plaintexts
28:	procedure SETPLAINTEXT(<i>p</i>)
29:	if <i>sender</i> exist and <i>flag</i> then
30:	if $hash(p)$ equals to $map(sender).hw$ then
31:	$i \leftarrow 0$
32:	for each delimiter d in p do
33:	if $ans[i]$ is equals to $p.split(d)$ then
34:	$correct \leftarrow true$
35:	else
36:	$correct \leftarrow false$
37:	map(sender).result.push(correct)
38:	$i \leftarrow i + 1$
39:	
40:	// Get the result of correction
41:	procedure GETRESULT()
42:	if <i>sender</i> exist and <i>flag</i> then
43:	return $map(sender).result$

5.3. A smart contract to support collaborative scoring

To make scoring of open-ended questions more convincing by allowing the answers be judged by different scorers, a smart contract is designated to implement a collaborative scoring framework. This smart contract consists of seven basic functions (Figure 12): (1) submit ciphertext, (2) start grading, (3) submit plaintext, (4) register scorer, (5) get assignment, (6) score, and (7) get result. Once again, students need to upload their ciphertexts using function (1) as the commitments, submit the assignment answer plaintexts by calling function (3), and after the deadline or after the teacher starts the correction/scoring process applying function (2). Notice that function (2) plays only the role of locking and unlocking functions, as listed in Table 3, without asking for standard procedures of scoring to ensure scorers following their own opinions. Function (4) allows the teacher (contract owner) to add scorers to the smart contract at any moment and the scorers can then apply function (5) to see student's information (ciphertext, plaintext and address) they need for scoring the submissions. The scoring results using function (6), whereas the students and the teacher can find the scorers and a weighted final score.

This smart contract design provides an efficient method for collaborative scoring. It helps manage tasks and integrate information into one simple platform while still ensures all scorers' and students' anonymity so that each judgement can be made without interference by other factors. With this smart contract, scores are given trustworthily and faithfully so that the final scores can reflect the true learning outcome and thus the certificate of the course or the achievements accomplished in the course can be more convincing. The algorithm is shown in Figure 13.



Figure 12. Schematic diagram of the proposed smart contract based autonomous collaborative scoring system



)		
Functions	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Deployment	En	En	Dis	En	Dis	Dis	Dis
After (2) is called	Dis	Dis	En	En	En	En	En
Note	(1): Sub	mit cipher	text		(.	5): Get assig	gnment
	(2): Start scoring (6): Score						
	(3): Sub	mit plainte	ext		()	7): Get resu	lt
	(4): Reg	gister score	er				

Note. En: Enabled, Dis: Disabled.

Figure	12	Algorithma	Callaborativa	acomina
г igure	15.	Algorithm:	Conaborative	scoring

1: 2: 3: 4:	// Teacher creates Contract procedure CONSTRUCTOR() $owner \leftarrow sender$ $flag \leftarrow false$	28: 29: 30: 31:	<pre>// Teacher adds scorers procedure SETSCORERS(address) if sender is owner and address is not exist then scorer.push(address)</pre>
5: 6: 7: 8: 9:	<pre>// Students upload ciphertexts before deadline procedure SETSTUDENTS(ciphertext) if flag then return if sender already exist then</pre>	32: 33: 34: 35: 36: 37: 38: 39: 40:	$\label{eq:constraint} \begin{array}{l} // \mbox{ Scorers find assignment to correct} \\ \mbox{procedure GETASSIGNMENTTOCORRECT}(address) \\ \mbox{if sender is scorer and } address is student and $flag$ then $a \leftarrow map(address).addr$ $c \leftarrow map(address).hw$ $p \leftarrow map(address).hw$ $p \leftarrow map(address).plaintext$ $return (a, c, p) $} \end{array}$
 11: 12: 13: 14: 15: 16: 	$map(sender).hw \leftarrow ciphertext$ else $newStudent \leftarrow map(sender)$ $newStudent.addr \leftarrow sender$ $newStudent.hw \leftarrow ciphertext$ $addresses.push(sender)$	40: 41: 42: 43: 44: 45: 46: 46: 47: 48:	// Scorers give scores procedure SETSCORE(addr,s) if sender is scorer and addr is student and flag then $J \leftarrow scorer.length()$ $S \leftarrow map(addr).score.length()$ if S is not equal to J then for i from 1 to $J - S$ do map(addr).score.push(0)
 17: 18: 19: 20: 21: 22: 23: 	<pre>// Teacher starts correction work procedure SETSTARTCORRECT() if sender is not owner or flag then return flag ← true // Students upload plaintexts</pre>	49: 50: 51: 52: 53: 54: 55: 55: 56: 57:	$n \leftarrow sender \text{ position in } scorer$ $map(addr).score[n] \leftarrow s$ // Get results procedure GETRESULT(address) if sender is exist and flag then $s \leftarrow map(address).score$ return s
24: 25:	procedure SETPLAINTEXT(<i>plaintext</i>) if <i>sender</i> exist and <i>flag</i> then		

26: $map(sender).plantext \leftarrow plaintext$ 27:

6. Experiment

6.1. Implementation

The purposed work is realized on the Ethereum blockchain network with designed application tools to integrate all the mechanisms introduced in Section 4 and 5. For simplicity and re-producibility, the proposed blockchain system is built based on the Ethereum source code (https://github.com/ethereum/goethereum), programmed in Go language. The application tools are important keys to make the realized assignment scoring system much more user friendly. They cover all complex procedures for the users (students and teachers) so that everyone can use the system with ease by few simple selections without the need to understand the principles and theories of blockchain beforehand, which is in fact a desired scenario in real usage.

The application tools include three main modules: the cryptography module, the blockchain module, and the student identity module. Two versions of the application tools are created: one for the students and the other for the teachers. To summarize the tools, we used Nodejs for blockchain interactions, Go for Chaotic Cryptography (Amigo et al., 2007) and Python for user-interface, RSA (Shand & Vuillemin, 1993) and Secret Sharing (Shamir, 1979) for encryption/decryption. The procedures and the user interface of the application are illustrated in Figure 14 and Figure 15, respectively.





Figure 15. The application user interface for teachers (left) and students (right)



6.2. Performance test: randomness of the chaotic random number

In our work, a chaotic map based random number generating module is used to hash the inputted plaintexts for protecting the students' anonymity. Thus, the security of the Chaotic Cryptography module is directly correlated to the randomness of the generated random numbers. The experiment results show the high randomness of the generated random numbers and comparing the randomness between two generated results with two seeds differed in a tiny difference.



Figure 16. The noisy image generated with the seed value 12345678





Figure 18. The difference image between the two noisy images given in Figure 14 and Figure 15



Figure 16 (left) shows the 512×512 noise image corresponding to the generated random numbers, with the given seed 12345678 and the associated histogram (Figure 16 (right)) verifies that their corresponding distribution is very close to the uniform one. Figure 17 (left) shows another 512×512 noise image which is created with the seed value of 12345677 and as shown in Figure 17 (right) the resultant histogram is once again very close to uniform distribution. Although there is only a single digit difference between the two seeds, the comparison given in Figure 18 shows that there is a huge amount of pixel changes (99.6%) between the two images. Therefore, the outcomes of the adopted chaotic random number generating module are highly unpredictable and will bring large benefits to students' privacy and security.

6.3. Comparison with related work

Comparing our work with a popular centralized online course platform (https://ceiba.ntu.edu.tw/) called CEIBA used at first author's university, our system has advantages in system transparency and fairness to students (as shown in Table 4). In the centralized platform, information can be uploaded or deleted without being recorded, that is, those actions cannot be tracked by involved members (mostly students). Thus, students or teachers could miss some deleted information and result in dissensions. Additionally, if students use their true identities to interact with their teachers, this may result in teachers treating each student unequally due to an implicit stereotype. Therefore, using blockchain properties to openly track every information and activity will make the system much more *transparent* to avoid lots of unnecessary disputes between students and teachers by treating every on-chain member *equally*. Those smart contracts introduced in Section 5 not only help teachers distribute some heavy workloads but also ensure better *fairness* to every student. Moreover, the crypto system used in this work can prevent students from cheating.

Moreover, most of previous research focused on managing post stage of educational activities, for example, recording and sharing students' certificates and degrees between colleges to protect the "results" of assessment processes. In contrast, our work realized a design that is aiming at managing information security to *protect the "procedure" of assessment*, including assignment submissions and scoring, to ensure the *transparency* and *fairness* of educational assessment.

Table 4. Com	parison of the	purposed w	ork and the	traditional of	online course	platform. (CEIBA
10000 1.0011	surround of the	parposea n	one and the	in a controllar (omme course	pracioni, v	

	The purposed work	Traditional educational online platform
Decentralized	Decentralized	Centralized
Transparency	Blockchain property	Centralized
Fairness	Smart Contract, Blockchain property	Depend on teachers
Prevent cheating	Cryptography, Blockchain property	-
Speed	Latency due to Encryption/Decryption	-
Data Preservation	Maintained by all members	Centralized
Liveness, Stability	Depend on all members	Depend on administrator

7. Discussion

The design of the proposed system focuses mostly on realization of a transparent and fair assignment scoring platform based on the blockchain technology. Both *system performance (security and stability)* and *pedagogical feasibility* are considered:

Regarding the *performance of the proposed system*, the *security* properties described in this work was tested and proved to be stable (low latency without error) using 20 nodes equipped with a 2.80GHz CPU and 16GB RAM. The *stability* of the blockchain seems to be able to reliably handle lots of users because it is Ethereum-based blockchain, which has been used by millions of users. Besides, our system uses the crypto-hash function and the RSA algorithm, whose security depends on the hash function and the key size. National Institute of Standards and Technology (NIST) provided the estimated maximum-security (Barker & Dang, 2016) and the 1024-bit RSA we used can achieve 80-bits security strength which keeps an acceptable trade-off between security and encoding speed. In the future, longer key length could be used when computing power improves to the increase security level. Moreover, a larger scale real test should be performed before the system is ready to be deployed in practical usage. To increase the practical value of our current system, the user interface should be more user-friendly. Some procedures of the proposed system can also be automated to make scoring more effective. For example, automatically decrypting ciphertext once the teacher's account received a certain amount of submissions from students.

Regarding the *pedagogical feasibility*, the proposed system provides many features that are aligned with the emerging trends in education. The "*peer evaluation*" smart contract relies on the submission order of students' submitted assignments and could be exploited when a group of students conspired to upload their assignments (in ciphertext form) at the same time. Using hash functions to generate random numbers can make the exchanging behavior more unpredictable. But, again, this may result in too high of computational cost when there are too many students enrolled in the course. Besides, the random numbers are predictable by those who decide the seeds of the hash functions (e.g., the responsible teachers) or those who generate the blocks (i.e., the miners). Additionally, blockchain provides a more *student-centered environment*, students have an easy way to store and manage their portfolio, projects, credits, and degrees, which contributes to self-directed learning. The

system also allows educators, universities, and institutes to manage student-related affairs, share their information with other universities, and track students' learning histories and outcome. It can also prevent improper activities, such as cheatings or forgeries, with the aid of blockchain, a decentralized and transparent system where every activity can be verified and supervised by all involved members. With blockchain, a student can apply for the entrances to colleges without printing mass of diplomas or certificates of programs learnt; instead, colleges can find student's information. This will not only save resources and time, but also establish fairness, transparency, and security of information flow.

Although it is expected to establish an efficient way for supervising scoring-related activities and ensuring fairness to all members, the latency caused by the involved encryption processes becomes the major obstacle to its adaption in practice. The most obvious latency is caused by the RSA module, which takes approximately one minute to encrypt a plaintext with just 100 words. Fortunately, this comes from the huge time cost from programming implementation, which can be solved by optimization techniques.

In the future, combining our work with other related works to integrate the merits of blockchain technology into higher level education usage, such as sharing and maintaining students' certificate and learning results between institutions and colleges, is of great interest. Although this goal is currently difficult to achieve since it requires the support of cross-chain techniques. However, the cross-chain system integration might build a complete blockchain-based educational system, from information sharing between institutions, basic interactions between teachers and students, to establish a true transparent and fair educational system for all students, teachers, and administration staffs.

8. Conclusions

This paper presents a design of blockchain-based assignment scoring mechanism for online learning. Our goal is to take advantages of blockchain properties and cryptography algorithms to build a transparent and secure teacher-student interaction system for online assessment. The fairness of scoring can be guaranteed by anonymity of the proposed blockchain architecture and the collaborative scoring policy. Although the online learning system will pay extra computational cost and related administrative procedures need to be made to use the proposed scoring mechanisms, our work is one initial step in designing and developing a feasible scoring mechanism to achieve fairer and more secure assessment for the rapid-growing online learning. The trend of education is moving toward online model. The proposed methodology can contribute to the high-quality assessment for online learning. In the future, empirical studies could be conducted by embedding the proposed mechanism in a real online learning platform such that its effectiveness in real educational applications could be examined. In addition, big data solutions and the architecture design (e.g., using proof-of-stake to reduce computational power and carbon footprint) can also be considered to enhance the feasibility of the proposed scoring mechanism. More advanced algorithms can also be applied to improve the performance of the proposed scoring system. For example, picking another random number generator with higher randomness and efficiency with less computational cost is an important task for smart contract designer to provide more unpredictability, and thus, achieving real fairness for all members.

References

Amigo, J. M., Kocarev, L., & Szczepanski, J. (2007). Theory and practice of chaotic cryptography. *Physics Letters A*, 366(3), 211-216. http://doi.org/10.1016/j.physleta.2007.02.021

Apampa, K. M., Wills, G., & Argles, D. (2010). User security issues in summative e-assessment security. *International Journal of Digital Society (IJDS)*, 1(2), 1-13. http://doi.org/10.20533/ijds.2040.2570.2010.0018

Barker, E., & Dang, Q. (2016). *NIST special publication 800-57 part 1, revision 4. Recommendation for Key Management Part 1: General.* National Institute of Standards and Technology Special Publication (NIST). http://dx.doi.org/10.6028/NIST.SP.800-57pt1r4

Caballé, S., Miguel, J., Xhafa, F., Capuano, N., & Conesa, J. (2017). Using trustworthy web services for secure e-assessment in collaborative learning grids. *International Journal of Web and Grid Services*, *13*(1), 49-74. http://doi.org/10.1504/IJWGS.2017.082059

Calderbank, M. (2007). *The RSA cryptosystem: History, algorithm, primes.* Math. Uchicago. Edu. https://www.math.uchicago.edu/~may/VIGRE/VIGRE2007/REUPapers/FINALAPP/Calderbank.pdf

Castella-Roca, J., Herrera-Joancomarti, J., & Dorca-Josa, A. (2006). A Secure e-exam management system. In *Proceedings of the First International Conference on Availability, Reliability and Security (ARES'06)*. IEEE. http://doi.org/10.1109/ARES.2006.14

Chen, G., Xu, B., Lu, M., & Chen, N. S. (2018). Exploring blockchain technology and its potential applications for education. *Smart Learning Environments*, 5(1), 1-10. http://doi.org/10.1186/s40561-017-0050-x

Dannen, C. (2017). Introducing Ethereum and solidity. Apress. http://doi.org/10.1007/978-1-4842-2535-6_4

Das, M., Tao, X., & Cheng, J. C. P. (2021). A Secure and distributed construction document management system using Blockchain. In Toledo Santos, E., & Scheer, S. (Eds.), *Proceedings of the 18th International Conference on Computing in Civil and Building Engineering* (pp. 850–862). https://doi.org/10.1007/978-3-030-51295-8_59

Grech, A., & Camilleri, A. (2017). *Blockchain for education*. Publications Office of the European Union. http://doi.org/10.2760/6064910.1007/s11191-017-9891-5

Hernandez-de-Menendez, M., Escobar Díaz, C., & Morales-Menendez, R. (2020). Technologies for the future of learning: State of the art. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 14(2), 683-695. http://doi.org/10.1007/s12008-019-00640-0

Hoy, M. B. (2017). An Introduction to the blockchain and its implications for libraries and medicine. *Medical Reference Services Quarterly*, 36(3), 273-279. http://doi.org/10.1080/02763869.2017.1332261

Jain, N., Agrawal, T., Goyal, P., & Hassija, V. (2019). A Blockchain-based distributed network for secure credit scoring. In 2019 5th International Conference on Signal Processing, Computing and Control (ISPCC) (pp. 306-312). IEEE. http://doi.org/10.1109/ISPCC48220.2019.8988510

Jansen, M., Hdhili, F., Gouiaa, R., & Qasem, Z. (2019). Do smart contract languages need to be Turing complete? In *International Congress on Blockchain and Applications* (pp. 19-26). Springer, Cham. http://doi.org/10.1007/978-3-030-23813-1_3

Kocarev, L., & Lian, S. (Eds.). (2011). Chaos-based cryptography: Theory, algorithms and applications (Vol. 354). Springer Science & Business Media. http://doi.org/10.1007/978-3-642-20542-2

Lam, T. Y., & Dongol, B. (2020). A Blockchain-enabled e-learning platform. *Interactive Learning Environments*, 1-23. http://doi.org/10.1080/10494820.2020.1716022

Li, C., Guo, J., Zhang, G., Wang, Y., Sun, Y., & Bie, R. (2019). A Blockchain system for E-learning assessment and certification. In 2019 IEEE International Conference on Smart Internet of Things (SmartIoT) (pp. 212-219). IEEE. http://doi.org/10.1109/SmartIoT.2019.00040

Lilley, M., Meere, J., & Barker, T. (2016). Remote live invigilation: A Pilot study. Journal of Interactive Media in Education, 2016(1). http://doi.org/10.5334/jime.408

Nakamoto, S. (2008). Bitcoin: A Peer-to-peer electronic cash system. Bitcoin. https://bitcoin.org/bitcoin.pdf

Nazare J., Duffy K., & Schmidt J. P. (2016). What we learned from designing an academic certificates system on the blockchain. MIT Media Lab. https://medium.com/mit-media-lab/what-we-learned-from-designing-an-academic-certificates-system-on-the-blockchain-34ba5874f196

Ngqondi, T., Maoneke, P. B., & Mauwa, H. (2021). A Secure online exams conceptual framework for South African universities. *Social Sciences & Humanities Open*, *3*(1), 100132. http://doi.org/10.1016/j.ssaho.2021.100132

Pavlov, O. V., & Katsamakas, E. (2021). COVID-19 and financial sustainability of academic institutions. *Sustainability*, *13*(7), 3903. http://doi.org/10.3390/su13073903

Porter, S. (2015). *To MOOC or Not to MOOC: How can online learning help to build the future of higher education?* Chandos Publishing. http://doi.org/10.1080/00049670.2016.1183469

Raimundo, R., & Rosário, A. (2021). Blockchain system in the higher education. *European Journal of Investigation in Health, Psychology and Education, 11*(1), 276-293. http://doi.org/10.3390/ejihpe11010021

Rivest, R. L., Shamir, A., & Adleman, L. (1978). A Method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2), 120-126. http://doi.org/10.1145/359340.359342

Sabbah, Y. W. (2017). Security of online examinations. In *Data Analytics and Decision Support for Cybersecurity* (pp. 157-200). Springer. http://doi.org/10.1007/978-3-319-59439-2_6

Shamir, A. (1979). How to share a secret. *Communications of the ACM*, 22(11), 612-613. http://doi.org/10.1145/359168.359176

Shand, M., & Vuillemin, J. (1993). Fast implementations of RSA cryptography. In *Proceedings of IEEE 11th Symposium on Computer Arithmetic* (pp. 252-259). IEEE. http://doi.org/10.1109/ARITH.1993.378085

Sharples, M., & Domingue, J. (2016). The Blockchain and kudos: A Distributed system for educational record, reputation and reward. In *European conference on technology enhanced learning* (pp. 490-496). Springer, Cham. http://doi.org/10.1007/978-3-319-45153-4_48

Skiba, D. J. (2017). The Potential of blockchain in education and health care. *Nursing education perspectives*, *38*(4), 220-221. http://doi.org/10.1097/01.NEP.00000000000190

Sudaryono, S., Aini, Q., Lutfiani, N., Hanafi, F., & Rahardja, U. (2020). Application of blockchain technology for iLearning student assessment. *Indonesian Journal of Computing and Cybernetics Systems (IJCCS)*, 14(2), 209-218. http://doi.org/10.22146/ijccs.53109

Traoré, I., Nakkabi, Y., Saad, S., Sayed, B., Ardigo, J. D., & de Faria Quinan, P. M. (2017). Ensuring online exam integrity through continuous biometric authentication. In *Information Security Practices* (pp. 73-81). Springer, Cham. http://doi.org/10.1007/978-3-319-48947-6_6

Turkanović, M., Hölbl, M., Košič, K., Heričko, M., & Kamišalić, A. (2018). EduCTX: A Blockchain-based higher education credit platform. *IEEE access*, 6, 5112-5127. http://doi.org/10.1109/ACCESS.2018.2789929

Urosevic, K. A. (2019). *Student authentication framework for online exams outside of school* (Unpublished master thesis). Laurea University of Applied Sciences, Vantaa, Finland. https://urn.fi/URN:NBN:fi:amk-201902061979

Wright, R. D. (Ed.). (2014). Student-teacher interaction in online learning environments. IGI Global. http://doi.org/10.4018/978-1-4666-6461-6