Journal of

Educational Technology & Society

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

Volume 24 Issue 2
April 2021
ISSN: 1436-4522 (online)
ISSN: 1176-3647 (print)
http://www.j-ets.net/

Apr. 2021 Journal of Educational Technology & Society
vol.24
no.2

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

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

Educational Technology & Society
An International Journal

Aims and Scope

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

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

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

Publisher
International Forum of Educational Technology & Society

Host
National Yunlin University of Science and Technology, Taiwan

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

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

Advertisements

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

Abstracting and Indexing

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

Guidelines for authors

Submissions are invited in the following categories:

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

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

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

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

Submission procedure

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

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

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

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

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

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

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

Educational Technology & Society

Volume 24 Number 2 2021

Table of contents

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

Yang, C. C. Y., Chen, I. Y. L., Akçapınar, G., Flanagan, B., & Ogata, H. (2021). Using a Summarized Lecture Material Recommendation System to Enhance Students' Preclass Preparation in a Flipped Classroom. *Educational Technology & Society*, *24* (2), 1–13.

Using a Summarized Lecture Material Recommendation System to Enhance Students' Preclass Preparation in a Flipped Classroom

Christopher C. Y. Yang[1*], Irene Y. L. Chen[2], Gökhan Akçapınar[3], Brendan Flanagan[4] and Hiroaki Ogata[4]

[1]Graduate School of Informatics, Kyoto University, Japan // [2]Department of Accounting, National Changhua University of Education, Taiwan // [3]Department of Computer Education and Instructional Technology, Hacettepe University, Turkey // [4]Academic Center for Computing and Media Studies, Kyoto University, Japan // yang.yuan.57e@st.kyoto-u.ac.jp // irene@cc.ncue.edu.tw // gokhanakcapinar@gmail.com // flanagan.brendanjohn.4n@kyoto-u.ac.jp // hiroaki.ogata@gmail.com

[*]Corresponding author

(Submitted May 27, 2020; Revised June 10, 2020; Accepted December 25, 2020)

ABSTRACT: Research has revealed the positive effects of flipped classroom approaches on students' learning engagement and performance compared with conventional lecture-based classrooms. However, because of a lack of out-of-class learning support, many students fail to comprehensively prepare the provided lecture materials before class. One promising solution to this problem is recommendation systems in the educational area, which have been instrumental in helping learners identify useful and relevant lecture materials that satisfy their learning needs. Thus, in this study, we propose a summarized lecture material recommendation system, which is integrated into an e-book reading system as an enhancement of the flipped classroom approach. This system helps students identify pages that contain essential knowledge that must be thoroughly studied before class. The proposed system was constructed on the basis of our previous work. In this study, a quasi-experiment was conducted in a graduate course that implemented the flipped classroom model: experimental group students learned with the proposed system, whereas the control group students had no access to the additional features. The findings of this study suggest that students who learn with the proposed recommendation system significantly outperform those who learn without the system in a flipped classroom in terms of their learning outcomes and engagement in preclass preparation.

Keywords: Recommendation systems, Flipped classrooms, Preclass preparation, Learning analytics, Learning outcomes

1. Introduction

Researchers have revealed the positive effects of flipped classrooms on students' motivation and performance compared with conventional lecture-based classrooms (Lai & Hwang, 2016). As defined by Bishop and Verleger (2013), the flipped classroom is a technology-supported pedagogical approach that combines asynchronous computer-assisted instruction outside the classroom and interactive in-class learning activities. Flipped classrooms are considered an effective model for engaging students in active learning and involving meaningful teaching and learning strategies during the in-class and out-of-class learning process (Forsey, Low, & Glance, 2013).

However, few studies have investigated how to improve students' out-of-class learning instruction in the flipped classroom model (Jensen et al., 2018). For flipped classroom approaches to be successful, educators must have confidence that their students learned the necessary information and skills from the lecture materials before class. Without appropriate guidance or support, most students tend to present a low level of learning and responsibility during the learning process in a flipped classroom (McLaughlin et al., 2013; Sun, Wu, & Lee, 2017). This lack of clarity regarding optimal class preparation instructions and methods may explain why educators are hesitant to adopt flipped teaching (Jensen et al., 2018). In addition, students' learning achievements in the course have also been reported to strongly correlate with their learning engagement (Yang et al., 2020; Kahu, 2013) and frequently used to examine the effects of their learning approaches on students' learning in flipped classrooms (e.g., Jovanović et al., 2017; Lai & Hwang, 2016). These reports emphasize the importance of considering students' behavioral engagements as measurements when evaluating the effects of educational systems on students' learning in flipped classrooms.

Shimada et al. (2017) revealed that learners are not always willing to prepare for classes, particularly when the volume of the lecture material is too large for them to adequately prepare before class. In their work, it is reported that students have a marked preference for studying summarized lecture material instead of the original material or additional content from external learning environments when aiming for thorough preparation before class. Accordingly, they recommended that students be provided with summarized lecture material instead of requiring them to study the full content; such a measure can increase students' preclass browsing time and overall learning performance.

Learning has shifted from the conventional classroom to the e-learning environment, which uses electronic media, principally through the Internet (George & Lal, 2019). With increasing research attention on recommendation approaches in e-book-based learning resource as well as the rapidly growing amount of online lecture materials, recommendation systems (RSs) in technology-enhanced learning (TEL) have become valuable in supporting learners in identifying useful and relevant lecture materials that satisfy their learning needs (Tarus, Niu, & Mustafa, 2018). The aim of developing an educational RS is to reduce information overload by retrieving the most relevant information and services from a large amount of data (Lu et al., 2015). Thus, current intelligent agents and RSs have been widely studied and accepted as solutions for overcoming information retrieval challenges by learners because of information overload (Montaner, López, & De La Rosa, 2003). These systems could therefore address some of the concerns in flipped classroom models, such as students not preparing before class (Louhab et al., 2020).

However, most existing and proposed educational RSs focus on recommending lecture materials from an external learning environment (Bauman & Tuzhilin, 2018; Ghauth & Abdullah, 2010). Thus, comprehensively satisfying students' learning needs, particularly for students' preclass preparation in flipped classrooms, remains challenging. Therefore, investigations regarding recommendations within the lecture material itself is quite rare, and further clarification is needed.

To address the aforementioned concerns regarding flipped classrooms and improve students' preclass preparation, in this study, based on our previous work (Yang et al., 2019), we propose the summarized lecture material recommendation system that utilizes several text processing and image processing techniques to generate recommendations within the lecture material itself. The proposed system is considered an enhancement of the flipped classroom model as it aims to retrieve the most important information from an expected large amount of contents from the online learning materials and thereby providing suitable guidance to support students' preclass preparation in flipped classrooms.

In contrast to previous works on flipped classroom approaches and educational RSs, internal content (e-book pages) is the target in this proposed system. Instead of the entire digital lecture material, summarized lecture material is generated by collecting the recommended pages from the original lecture material. Moreover, a supervised machine learning approach is applied that extracts more features from e-book text and image content compared with the approach presented in Shimada et al. (2017). To elucidate the effects of the proposed system on students' learning outcomes and engagement in preclass preparation, we investigated the following research questions:

RQ1: Do students who learn with the proposed system achieve better learning outcomes in comparison with students who learn without the system in a flipped classroom?
RQ2: Do students who learn with the proposed system exhibit greater engagement in preclass preparation in comparison with students who learn without the system in a flipped classroom?

2. Literature review

2.1. Flipped classrooms

Researchers have continually investigated how to maximize students' academic performance in higher education. The flipped classroom is gaining popularity as a means to support student learning in such contexts (Boevé et al., 2017). The organization of flipped classroom lectures requires students to prepare for the in-class meetings, often facilitated through additional online video lectures, and demands involvement during lectures by means of problem-solving and peer instruction (Abeysekera & Dawson, 2015).

The flipped classroom uses a student-centered approach, focusing on student learning and placing the responsibility for learning more on students than on teachers while encouraging them to experiment (Sams, 2011). However, researchers have argued that without appropriate guidance or support, most students tend to exhibit a low level of learning and responsibility during the learning process in a flipped classroom (McLaughlin et al., 2013; Sun, Wu, & Lee, 2017). Thus, to adequately support students' preclass preparation in flipped classroom models, Diwanji, Hinkelmann, and Witschel (2018) proposed an application based on artificial intelligence (AI) and data analysis techniques to intrinsically motivate the students' preclass preparation in a flipped classroom. Louhab et al. (2020) proposed the smart adaptive management for flipped learning plugin that can be combined with a Learning Management System to provide students with adapted lecture materials according to their knowledge level and skills when learning in a flipped classroom. Their findings indicated that the students who used the plugin obtained more favorable final examination results than students who did not. This also exhibits the positive effects on students' knowledge levels when integrating the advanced educational technologies into flipped classroom settings.

The aforementioned studies have demonstrated the considerable advantages of implementing flipped classrooms as a new teaching strategy. However, few studies have focused on applying educational RSs that can provide suitable guidance to support students' preclass preparation in flipped classrooms. Hence, investigating the effects of educational RSs on students' learning outcomes and their engagement in preclass preparation in a flipped classroom (i.e., reading time) is paramount.

2.2. Educational recommendation systems

Using a different educational setting and novel methods, Bousbahi and Chorfi (2015) proposed an architecture for personalizing a massive open online course by using a case-based reasoning RS. Benhamdi, Babouri, and Chiky (2017) developed a new multi-personalized recommender system that considers a student's level of study, preference, and memory capacity when providing recommendations. Klašnja-Milićević et al. (2011) proposed an online tutoring system, Protus, in which learners are clustered on the basis of their learning styles. Furthermore, the learners' sequential patterns were mined. Finally, recommendations were generated by applying the collaborative filtering approach.

AI planning and case-based planning can also be applied to achieve a more personalized e-learning solution for students (Garrido, Morales, & Serina, 2016). Obeid et al. (2018) generated personalized recommendations by creating ontologies for students, employees, and higher educational institutions. Machine learning algorithms were then applied for clustering similar graduate students. Hussain et al. (2019) developed a system using machine learning algorithms such as support vector machine, logistic regression, naive Bayes classifier, and artificial neural networks. The objective was to predict students' difficulty when undertaking courses online and provide feedback to teachers on those who are using the system.

Using the features of educational RSs, the aforementioned systems were helpful in supporting students' preclass preparation because they could retrieve essential information from the overwhelming amount of online resources in different educational settings. However, most existing educational RSs were not designed to provide suitable guidance to support students' preclass preparation in flipped classrooms. To achieve such a goal, the proposed system takes into the advantages of AI techniques to generate recommendations within the online lecture material itself. Specifically, the proposed system predicts e-book pages that contain essential knowledge from online lecture materials as the learning guidance for students to perform a thorough preclass preparation in flipped classrooms.

3. A summarized lecture material recommendation system

3.1. System overview

The proposed system aims to rank e-book pages according to their corresponding importance using a machine learning approach. The proposed system selects the e-book pages that are top-ranked for containing essential information that must be learned in advance to form the summarized lecture material. Figure 1 illustrates the procedures of the proposed system. First, we collected different types of data from the lecture material uploaded to an e-book system as the system's input data. Second, we extracted text features and image features from the collected input data. Third, we preprocessed the extracted features. Fourth, we trained a machine learning model to retrieve class probabilities from the training process and rank the input e-book pages accordingly. Last, after selecting several top-ranked e-book pages in a preference range from the ranking, the summarized lecture material recommendation is generated as the output.

Because of the characteristics of machine learning-based systems, the proposed system becomes more predictive as it receives new materials from the same instructor and the same course. Notably, the proposed system only uses new lecture materials from the same instructor and course for model training. This is inspired by the notion that learning and educational technologies should be adopted and applied in course-specific contexts rather than relying on a generalized model that applies one trained prediction model in all the educational contexts (Gašević et al., 2016).

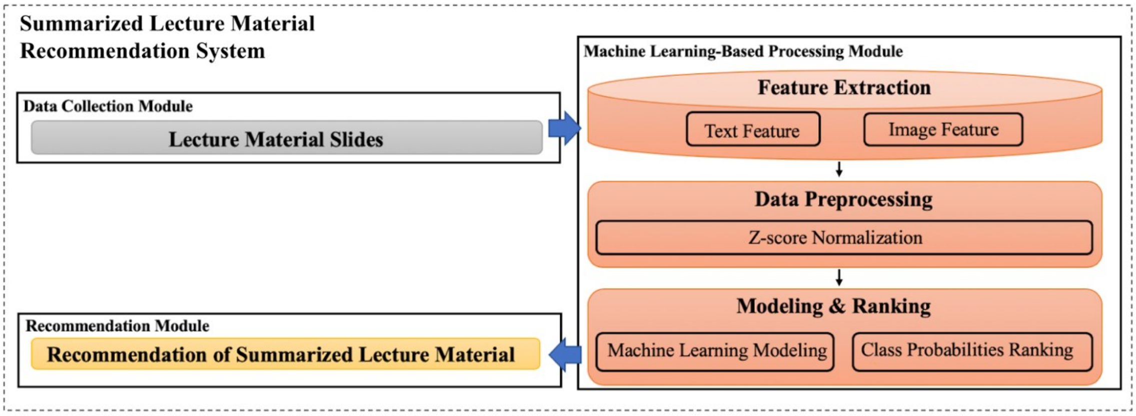


Figure 1. Procedures of the proposed system

3.2. Data collection, feature extraction, and data preprocessing

In this study, we used the BookRoll system (Ogata et al., 2015), which is an e-book reading system that allows users to browse the uploaded content anytime and anywhere. Figure 2 illustrates an example of BookRoll's user interface. Several functions such as page-turning, highlighting, bookmark making, note-taking, page jumping, and internal content searching are available. Students' reading behaviors while using BookRoll are stored in its database. During the data collection process, by using BookRoll, text content, and image content from the original digital lecture material were collected for further feature extraction and data preprocessing. Note that the types of BookRoll learning behavioral logs were described by the developers in their 2015 study (Ogata et al., 2015); therefore, they are not detailed in the present study.

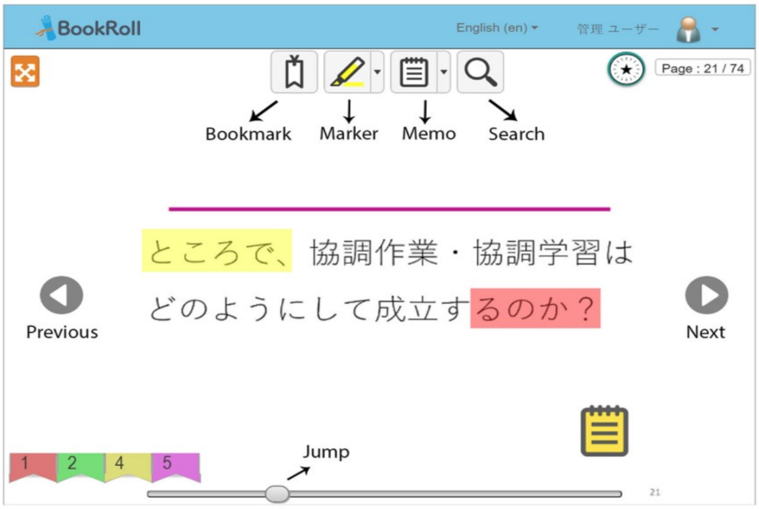


Figure 2. Screenshot of the BookRoll system's user interface

The characteristics of the essential document or e-book pages and the associated content features during the feature extraction process have been mentioned and applied in several studies (Neto, Freitas, & Kaestner, 2002; Shimada et al., 2017). Therefore, we extracted two categories of features from the collected data on the basis of the characteristics of essential pages, as described in Table 1, to train our summarized lecture material generator. In the present study, an essential document page or e-book page should have the following characteristics:

- Sufficient content to be worth browsing
- Unique visual content
- Keywords that appear frequently on the page

- Keywords that rarely appear throughout the whole lecture material
- Similarity to the title of the lecture material
- Similarity to other pages under the same lecture title

We applied text processing techniques including N-gram tokenizer, cosine similarity, and lemmatization to extract text features from the collected e-book pages. We used cosine similarity measurement to calculate text similarity between paired e-book pages and similarity to the title and keywords of the lecture, which represents the values for Page–Page cohesion, similarity to the title, and similarity to keywords, respectively. We counted the total number of characters and punctuations on a page from the generated corpus to represent the values of TotalChar and Punctuation. We applied the vector space modeling technique term frequency–inverse document frequency (TFIDF) weighting to calculate the weight of each term on the page and calculated the average TFIDF value for each page, representing the value of AvgTFIDF.

For image feature extraction, we applied image processing algorithms including the background subtraction method and the interframe difference method, which have both been used for the extraction of visual information from the lecture materials, to extract values representing Background subtraction and Background subtraction + Inter-frame difference. The details of the extracted image features were described in a previous article (Shimada et al., 2017).

Next, to obtain an improved distribution of the dataset that contains all the feature values of each e-book page, we preprocessed each extracted feature by using the z-score normalization method to redefine and transform the range of feature values into a smaller and specific range.

Table 1. Description of the extracted features

Feature	Description of the feature
TotalChar	Total characters on the page
AvgTFIDF	Average TFIDF after preprocessing by the bigram tokenizer
Similarity to title	Cosine similarity to the title of the content
Similarity to keywords	Cosine similarity to the keywords of the content
Page–Page cohesion	Sum of cosine similarities to the remaining pages
Punctuation	Total instances of punctuations on the page
Background subtraction	Total amount of foreground pixels on the page
Background subtraction + Inter-frame difference	Absolute foreground pixel differences in comparison with the previous page and the next page (higher value is selected)

3.3. Model generation and recommendation of summarized lecture material

In this study, our problem had a binary classification; however, we considered class probabilities as an output instead of class labels. This provides the flexibility to rank e-book pages on the basis of their possibilities to be positively classified. Thus, the instructor can determine which volume of the summarized lecture material becomes the recommendation for preclass preparation. In the proposed system, we requested that instructors label "recommended pages" in lecture material to represent the gold standard. These labeled pages are recommended reading before class because they contain essential information. After receiving the labeled lecture material, we trained a supervised machine learning model to generate the class probabilities of each page. Next, these e-book pages were ranked by class probabilities, and the top-ranked pages were then selected to generate the summarized lecture material recommendation.

3.4. Preliminary evaluation for selecting the best-performing model and constructing the proposed system

To evaluate the predictive performance of the machine learning models, we conducted a preliminary evaluation using the extracted text and image features of lecture material containing 91 pages. The instructor labeled 45 out of 91 pages as recommended reading for preclass preparation. The remaining 46 pages were then considered "not recommended" pages, indicating optional reading before class. In the preliminary evaluation, we trained and compared eight well-known machine learning classification models from the Python (Pedregosa et al., 2011) libraries that involve different algorithms for class probability prediction during the model training process. We used all the extracted text features and image features for the model training, and we ranked the input e-book pages by the predicted class probabilities accordingly. After the ranking process, we selected the top-ranked e-book pages to form the recommendation for learners' preparation before class. All the prediction models were trained with default parameters.

The models' performance was evaluated using precision, recall, and area under the curve (AUC) of receiver operating characteristic (ROC) metrics. The precision and recall metrics were calculated according to the confusion matrix, and the AUC was calculated from the ROC curve. The AUC value ranges between 0 and 1, and an AUC value near 0.5 indicates that the classification process is similar to a random guess. The values of precision and recall are always the same in each model because the number of pages that the proposed system predicted as the recommended pages are always equivalent to the number the instructor labeled as recommended pages. Hence, the performance was validated using threefold cross-validation, meaning that our dataset was randomly divided into three disjoint subsets of equal size in a stratified manner (maintaining the original class distribution) and each model was executed three times. In each repetition, one of the three subsets was used as the testing set, and the other two were combined to form the training set. The preliminary results are presented in Table 2. The results revealed that the multilayer perceptron classification model (indicated in boldface in Table 2) was the optimal model with the greatest potential for use in the development of our proposed system. These results provided the data necessary to implement our experiment, which applied the proposed system on a graduate course that employed the flipped classroom model.

Table 2. Preliminary evaluation results of model performance

Model	Precision	Recall	AUC
DTNB	0.422	0.422	0.429
JRip	0.489	0.489	0.494
Random Forest	0.556	0.556	0.560
J48	0.533	0.533	0.538
BayesNet	0.422	0.422	0.429
Gaussian Naive Bayes	0.467	0.467	0.472
Logistic Regression	0.622	0.622	0.626
Multilayer Perceptron	**0.667**	**0.667**	**0.670**

3.5. Validation of the system's performance in comparison with human decisions

After the preliminary evaluation for finding the ideal model, we asked instructors to compare the recommendation provided by the system to the material they identified as essential reading before class. Tests were conducted on the six different lecture materials that were provided to students for preclass preparation in flipped classroom settings during the experiment. The results of the predictive performance of the system, as measured by precision, recall, and AUC, are shown in Table 3. The volume of the lecture materials (in pages) is also presented. The proposed system exhibited decent predictive performance, with scores higher than 60% for precision, recall, and AUC for each week.

Table 3. Predictive performance of the proposed system in comparison with human decisions

Lecture Material	Precision	Recall	AUC	Volume (# of Page)
Lecture material 1	0.742	0.737	0.739	30
Lecture material 2	0.744	0.741	0.742	27
Lecture material 3	0.816	0.806	0.812	27
Lecture material 4	0.733	0.733	0.733	31
Lecture material 5	0.647	0.645	0.646	30
Lecture material 6	0.62	0.613	0.616	32

4. Experiment

4.1. Participants and study contexts

A quasi-experiment was conducted in a flipped classroom graduate course, namely creative learning, in a university in northern Taiwan. A total of 30 graduate students from the department of computer science participated in the experiment. Students enrolled in the course were required to use BookRoll to study the given scientific articles during out-of-class learning time. During each class, students presented and discussed the content of the scientific articles with their classmates in addition to receiving some instruction from their teacher. Among the participants, 15 students were assigned as the experimental group and the remaining 15 students were assigned as the control group. Students in the experimental group were allowed to use BookRoll freely for their in-class and out-of-class learning with additional requirements to the uses of summarized lecture material recommendations in addition to the original lecture material for preclass preparation. By contrast, students in the control group used BookRoll for their learning but without receiving any of the recommendations from our proposed system throughout the experiment.

4.2. Measurements

A pretest and posttest were issued by the course instructor. The pretest aimed to evaluate the students' prior knowledge of the course. It consisted of 10 open-ended questions pertaining to data science and statistical analysis (e.g., what is text mining, what is t-test, how do you define low-skill readers), with a perfect score of 100. The posttest aimed to assess the students' learning outcomes in the course. It also consisted of 10 open-ended questions pertaining to data science and statistical analysis (e.g., what is the key feature of the text mining method, what does the t-test show and how to interpret the results, what is the main reason why low-skill readers use text-marking strategies), with a perfect score of 100. In addition, two experts from the department of computer science at the participating university were invited to ensure that the pretest and posttest were sufficient to evaluate the students' prior knowledge and learning outcomes in the course. Pearson's correlation between the two tests was 0.026, indicating a low correlation between the two tests. Moreover, the Kuder–Richardson Formula 20 of the posttest was 0.59, indicating acceptable internal consistency (Cortina, 1993).

The following three indicators were monitored throughout the experiment to measure students' learning behaviors (redesigned as "engagement" or "engagements" in the following sections) in preclass preparation when learning in the flipped classroom. It is noted that considering the cases that participants just open the lecture materials and quickly surf through the contents or do not pay attention to the contents, we pruned the outlier BookRoll activities and sessions. In particular, activities that were taken with too much (more than 10 minutes) or too few (less than three seconds) time were eliminated from the analysis of learning behaviors in the following sections. Moreover, the outlier sessions, in particular, overly long sessions (include more than 50 occurrences of activities) and overly short sessions (include less than three occurrences of activities) were eliminated from the analysis as well:

- Time (hour): Time spent in preclass preparation. The students' preclass time spent on learning materials has been indicated as a major factor that needs to be taken into account when utilizing flipped classroom models (Persky & Hogg, 2017). With its potential in indicating the students' learning engagements in flipped classrooms, in this experiment, students' preclass time spent on BookRoll is measured by summing up the reading time spent on each reading activity during preclass preparation. As shown in Figure 3, the timestamp of each reading activity can be recorded by the BookRoll system. For example, the subtracted value of the two timestamps of reading activities A and B indicates the time spent on reading activity A, etc.
- Activity: Number of BookRoll activities completed during preclass preparation. The students' preclass activities have been evident with the relevance to their learning outcomes in the course when learning in a flipped classroom where students with more preclass activities in a flipped classroom obtained higher course performance (Jovanović et al., 2017). With its potential in indicating the students' learning engagements in flipped classrooms, in this experiment, students' preclass reading activity on BookRoll is measured by summing up the occurrence of reading activities (as shown in Figure 3). For example, a total of four rows of BookRoll activities shown in Figure 3 represents the total of four occurrences of preclass reading activity. Among them, only three activities will be taken into account for the analysis of BookRoll learning behaviors as the time taken for the third activity is more than 10 minutes.
- Session: Number of reading sessions during preclass preparation. The students' preclass learning sessions have been reported to positively correlate with their learning outcomes in the course when learning in a flipped classroom where students with more preclass learning sessions obtained higher course performance (Jovanović et al., 2017). With its potential in indicating the students' learning engagements in flipped classrooms, in this experiment, students' preclass reading session on BookRoll is measured by summing up the occurrence of the reading activity OPEN as shown in Figure 3. For example, a total of one row of BookRoll activity OPEN, shown in Figure 3, represents the total of one preclass reading session. In this case, after pruning the outlier activities, the occurrence of activities in this session is three; therefore, this reading session will be taken into account for the analysis of BookRoll learning behaviors. On the other hand, if the occurrence of BookRoll activities in the session is more than 50 or less than three, this session will not be taken into account for the analysis.

An example of the learning log collected in BookRoll is illustrated in Figure 3.

User ID	Content ID	Operation Name	Operation Date
15910	ed645f3821e	OPEN	2019-12-20 10:03:52
15910	ed645f3821e	NEXT	2019-12-20 10:05:17
15910	ed645f3821e	ADD MEMO	2019-12-20 10:07:03
15910	ed645f3821e	NEXT	2019-12-20 11:27:14

Figure 3. Example of the BookRoll learning behavioral log collected in this study

4.3. Experimental procedures

Figure 4 illustrates the flow chart of the experiment. The duration of the experiment was eight weeks. The instructor released the original version of the lecture material to the BookRoll system one week before the class to enable students to prepare at their own pace and when and where they chose. In the first week of the experiment, the course syllabus and usage of the BookRoll system were introduced to all the students who were participating in the experiment. Moreover, the experimental (n = 15) and control groups (n = 15) completed the pretest examination to evaluate their prior knowledge of the course. Between the second and the seventh week of the experiment, the two groups prepared for class in different manners. Students in the experimental group were allowed to use BookRoll freely for their preclass preparation. Moreover, they were explicitly required to use the proposed system to view the summarized lecture material recommendations in addition to the original lecture material for preclass preparation. By contrast, students in the control group prepared without receiving any of the recommendations from our proposed system. Notably, both groups learned in a flipped classroom using BookRoll; the only difference between the two groups was the use of the proposed system to support preclass preparation. The length of the summarized lecture material that was provided to students in the experimental group depended on the instructor (in case of this study, 50% of the lecture materials were identified as essential reading before the class every week). Figure 5 illustrates the different BookRoll user interfaces for the control and experimental groups. For the experimental group, recommendations for summarized lecture materials were provided on the lower left side of the page (bookmarks), and students could browse the content by clicking the bookmarks. In the last week of the experiment, a posttest was conducted to evaluate both groups' learning outcomes in the course.

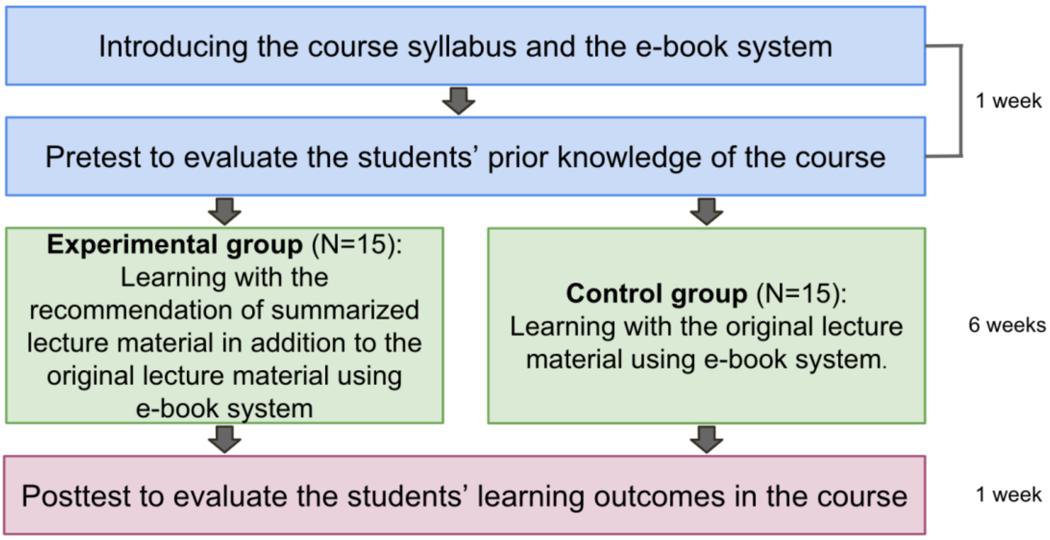


Figure 4. Experimental procedures

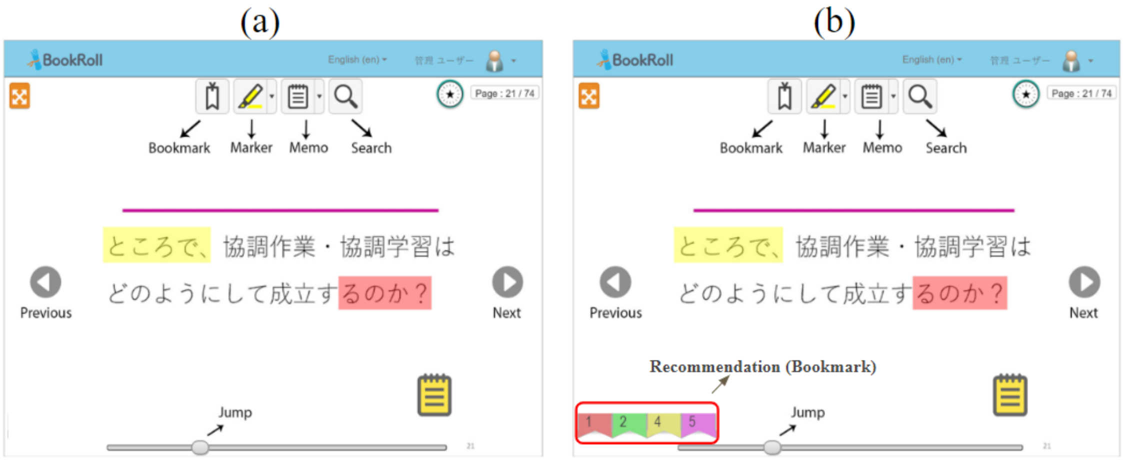


Figure 5. Different BookRoll user interfaces for (a) control group students that contained no recommendations and (b) experimental group students that contained recommendations on the lower left side using BookRoll's bookmark feature

5. Results

5.1. Analysis of learning outcomes

To investigate the effects of the proposed system on the learning outcomes of the students, the IBM SPSS one-way analysis of covariance (ANCOVA) was employed, with the students' pretest scores serving as the covariate. Levene's test of determining homogeneity of variance was not violated ($F = 0.68$, $p > .05$), suggesting that a common regression coefficient was appropriate for the implementation of the one-way ANCOVA. Moreover, the Shapiro–Wilk test was employed to examine the normality of the data. The results (control group: 0.93 [$p > .05$]; experimental group 0.95 [$p > .05$]) indicated that both groups' data were normally distributed in terms of posttest scores.

Table 4 presents the one-way ANCOVA results of the two groups' posttest scores. A significant difference was observed between the two groups of students in terms of their posttest scores ($F = 36.58$, $p < .001$, $\eta^2 = 0.58$), with an adjusted mean and adjusted standard deviation of 85.83 and 2.69, respectively, for the experimental group and 62.2 and 2.69, respectively, for the control group. The posttest scores of the experimental group were significantly higher than those of the control group. The effect size (η^2) was 0.58, indicating a medium to large effect size (Cohen, 1988). Overall, the results demonstrated that the proposed system exerted positive impacts on students' learning outcomes in flipped classrooms.

Table 4. One-way ANCOVA results of the posttest scores of the two groups

Group	*N*	Mean	*SD*	Adjusted mean	Adjusted *SD*	*F*	η^2
Experimental group	15	86.07	10.26	85.83	2.69	36.58^{***}	0.58
Control group	15	62.2	9.89	62.44	2.69		

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

5.2. Analysis of engagement in preclass preparation

Next, the effects of the proposed system on the two groups' engagement in preclass preparation were investigated. The Mann–Whitney U test was employed because the values of the three indicators did not meet the assumptions of data normality and homogeneity of variance, which are required for a parametric test.

Table 5 reveals the Mann–Whitney U test results of the three indicators of the students' engagement in preclass preparation. First, a significant difference was observed in terms of time spent on preclass reading by the median, 25th percentile, and 75th percentile of the experimental and control groups: 18.81 versus 4.12, 7.58 versus 2.5, and 30.65 versus 5.21, respectively ($U = 25$, $p < .001$). The preclass reading time of the experimental group students was significantly higher than that of the control group students. This indicates that students who used the proposed system had a greater engagement in terms of the time spent on preclass reading than did students who learned without the proposed system.

Table 5. Mann–Whitney U test results of the indicators of engagement in preclass preparation between two groups

Indicator	Group	*N*	Median	25th / 75 percentiles	*U*
Time (hour)	Experimental group	15	18.81	7.58 / 30.65	25^{***}
	Control group	15	4.12	2.5 / 5.21	
Activity	Experimental group	15	651	343 / 945	49^{**}
	Control group	15	209	157 / 325	
Session	Experimental group	15	35	30 / 66	59^{*}
	Control group	15	23	17 / 31	

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

A significant difference was observed in terms of the number of preclass learning activities engaged in by the median, 25th percentile, and 75th percentile of the experimental and control groups: 651 versus 209, 343 versus 157, and 945 versus 325, respectively ($U = 49$, $p < .01$). The number of preclass activities completed by students in the experimental group was significantly higher than that of students in the control group. This indicates that students who used the proposed system had greater engagement in terms of the number of preclass activities than did students who learned without the proposed system.

Moreover, a significant difference was evident in the number of preclass learning sessions engaged in by the median, 25th percentile, and 75th percentile of the experimental and control groups: 35 versus 23, 30 versus 17, and 66 versus 31, respectively ($U = 59$, $p < .05$). The number of preclass reading sessions of the experimental group was significantly higher than that of the control group. This indicates that students who used the proposed system had greater engagement in terms of the number of preclass reading sessions than did students who learned without the proposed system.

6. Discussion

6.1. Educational impacts of the proposed system on students' learning outcomes in flipped classrooms

The success of implementing flipped classroom approaches has been positively correlated with learners' learning achievements in the course based on the findings reported in Jovanović et al. (2017) and Lai and Hwang (2016). The experimental results in this study have revealed that the students who used the proposed system obtained significantly higher learning outcomes compared to the students who did not. These findings provide evidence that students who learn with the proposed system can achieve better learning outcomes than students who learn without it when learning with an e-book system in flipped classrooms. Specifically, it is implied that the students' learning outcomes in the course can be improved by helping them identify pages that contain essential knowledge that must be thoroughly studied before the class when learning with an e-book system in flipped classrooms. The results were consistent with previous findings reported in Shimada et al. (2017), where the students who learned with the summarized lecture material recommendations before class obtained more favorable learning outcomes than did students who did not have access to the recommendations. The results were also consistent with the findings of Louhab et al. (2020), where students who received guidance for preclass preparation through an RS exhibited higher knowledge levels than did students who did not use such a system. From the perspective of educational RSs, the results were consistent with those of previous studies (Bauman & Tuzhilin, 2018; Ghauth & Abdullah, 2010; Hsu, Hwang, & Chang, 2013), which have provided evidence that students who use an educational RS to support their learning tend to exhibit stronger academic performance than students who do not.

6.2. Educational impacts of the proposed system on students' preclass engagements in flipped classrooms

The importance of measuring students' behavior changes and engagements when evaluating the effects of newly proposed systems or methods on students' learning in flipped classroom settings has been emphasized based on the findings reported in Kahu (2013), Jovanović et al. (2017), and Lai and Hwang (2016). The experimental results in this study have revealed that the students who used the proposed system exhibited significantly greater engagement in preclass preparation compared to the students who did not. These findings provide evidence that students who learn with the proposed system exhibit greater engagement in preclass preparation than students who learn without it when learning with an e-book system in flipped classrooms. This also implies that by helping the students identify pages that contain essential knowledge that must be thoroughly studied before the class, the students tend to exhibit greater engagements in preclass preparation when learning with an e-book system in flipped classrooms. Notably, the current results are inconsistent with those of Shimada et al. (2017), where the students who learned with the summarized lecture material recommendations actually spent less time preparing for the lecture before class than did students who did not have access to the system. One reason for the disparity in the results could be the objective and design of the proposed system. In this study, the proposed system aimed to provide suitable guidance by recommending summarized lecture material in addition to the original lecture material to increase the students' engagements in preclass preparation, whereas Shimada et al. (2017) aimed to reduce the volume of the lecture material with the ultimate goal of minimizing the students' browsing time during preclass preparation; consequently, students were only allowed to browse the summarized lecture material. The divergent results of these two studies emphasize the importance of clarifying the objective of such systems (increasing or decreasing students' browsing time before class). The results obtained from the comparison results between the experimental group and the control group in terms of preclass reading time, number of learning activities, and number of learning sessions provide evidence that students who learned with the proposed system exhibited greater engagement in preclass preparation than did students who learned without it in a flipped classroom. The proposed system is expected to provide an extra option for instructors to automatically generate summarized lecture materials as the learning guidance for the learners' preclass preparation in flipped classrooms without expending too much time and effort. This also exhibits potential in enabling teachers to support students in increasing their behavioral engagements in preclass preparation in a more efficient manner.

6.3. Limitations and suggestions for future studies

This study had some limitations. Although the students' learning outcomes were normally distributed, only 30 students participated in the experiment. Therefore, future studies should employ a larger sample to obtain stronger results with greater external validity. Moreover, we investigated the effects of the proposed system using eight features from both text and image content in the e-book lecture material for the machine learning model training task. However, we did not establish whether other types of educational data can be used for similar recommendation tasks. Consequently, future studies could include other features from different educational data sources. Furthermore, although the proposed system was evident with its effects on students' learning outcomes in the course as well as their engagements in preclass preparation according to the experimental results, it is difficult to assert that the proposed system will not limit the students' strategies in preclass preparation when learning with flipped classroom settings. Therefore, future studies are suggested to investigate such an issue. Lastly, since we focused on the uses of the proposed system but did not include the investigation of the influence of participants' learning style or personality traits on their learning outcomes and engagement in preclass preparation, the analysis results might contain some bias. Consequently, it would be needed to extend our investigation with multi-modal approaches where participants' learning style or personality traits are combined with the analysis of data obtained from other sources such as self-reports or interviews with instructors, to better interpret the findings of this study, particularly in terms of the participants' engagements of learning.

7. Conclusions

Online learning continues to grow, in part, because of the reduced cost, increased flexibility regarding class schedule, and improved mobility when taking a class (Allen & Seaman, 2014). With a dramatically increasing volume of digital lecture material, methods to support students when they study the materials are more crucial than ever. These methods can streamline students' preclass preparation and thereby maximize their engagement in a flipped classroom. To support students in achieving better learning outcomes and exhibiting greater engagement in preclass preparation in flipped classrooms, on the basis of our previous work, a summarized lecture material recommendation system was proposed. The proposed system was integrated into an e-book system and implemented in a flipped classroom model, constituting a different learning approach than that employed in typical flipped classrooms.

In this study, a quasi-experiment was conducted in a graduate course in northern Taiwan that implemented the flipped classroom model to evaluate the effects of the proposed system on students' learning. The experimental group learned with the proposed system, whereas the control group had no access to the additional features. The findings of this study suggest that the proposed system can enhance the flipped classroom and mitigate, to some extent, some concerns with the flipped classroom approach, such as difficulty providing appropriate guidance to students for preclass preparation. The experimental students outperformed the control students in terms of their learning outcomes in the course and engagement in preclass preparation including reading time, number of learning activities, and number of learning sessions at statistically significant levels. This implies that students who learn with the proposed system can not only achieve better learning outcomes but also exhibit greater engagement in preclass preparation than can students who learn without the proposed system in a flipped classroom. Furthermore, the proposed system can be regarded as an effective option for instructors to automatically generate summarized lecture materials in addition to the original lecture materials without expending too much time and effort. This enables teachers to support students in enhanced flipped classrooms by increasing the students' engagements in preclass preparation and thereby improving their learning outcomes in the adopted flipped classroom approach.

As a follow-up to the current study, we are now developing a knowledge-based RS that can be adapted and personalized by considering the students' learning behavior, assessment data, and the knowledge structure of the lecture materials. We aim to develop and propose in the future a bigger framework for educational RSs for students' preclass preparation and postclass reflection by integrating the two RSs into the course.

Acknowledgement

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (B)20H01722, JSPS Grant-in-Aid for Scientific Research (S)16H06304 and NEDO Special Innovation Program on AI and Big Data 18102059-0.

References

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

Allen, I. E., & Seaman, J. (2014). *Grade change: Tracking online education in the United States*. Babson Park, MA: Babson Survey Research Group and Quahog Research Group.

Bauman, K., & Tuzhilin, A. (2018). Recommending remedial learning materials to students by filling their knowledge gaps. *MIS Quarterly*, *42*(1), 313-332.

Benhamdi, S., Babouri, A., & Chiky, R. (2017). Personalized recommender system for e-Learning environment. *Education and Information Technologies*, *22*(4), 1455-1477.

Bishop, J. L., & Verleger, M. A. (2013). The Flipped classroom: A Survey of the research. In *ASEE national conference proceedings, Atlanta, GA* (Vol. 30, No. 9, pp. 1-18).

Boevé, A. J., Meijer, R. R., Bosker, R. J., Vugteveen, J., Hoekstra, R., & Albers, C. J. (2017). Implementing the flipped classroom: An Exploration of study behaviour and student performance. *Higher Education*, *74*(6), 1015-1032.

Bousbahi, F., & Chorfi, H. (2015). MOOC-Rec: A Case based recommender system for MOOCs. *Procedia-Social and Behavioral Sciences*, *195*, 1813-1822.

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

Cortina, J. M. (1993). What is coefficient alpha? An Examination of theory and applications. *Journal of applied psychology*, *78*(1), 98-104. doi:10.1037/0021-9010.78.1.98

Diwanji, P., Hinkelmann, K., & Witschel, H. F. (2018). Enhance classroom preparation for flipped classroom using AI and analytics. In *Proceedings of the 20th International Conference on Enterprise Information Systems (ICEIS 2018)* (pp. 477-483). doi:10.5220/0006807604770483

Forsey, M., Low, M., & Glance, D. (2013). Flipping the sociology classroom: Towards a practice of online pedagogy. *Journal of Sociology*, *49*(4), 471-485.

Garrido, A., Morales, L., & Serina, I. (2016). On the use of case-based planning for e-learning personalization. *Expert Systems with Applications*, *60*, 1-15.

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The Effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68-84.

George, G., & Lal, A. M. (2019). Review of ontology-based recommender systems in e-learning. *Computers & Education*, *142*, 103642. doi:10.1016/j.compedu.2019.103642

Ghauth, K. I., & Abdullah, N. A. (2010). Learning materials recommendation using good learners' ratings and content-based filtering. *Educational technology research and development*, *58*(6), 711-727.

Hsu, C. K., Hwang, G. J., & Chang, C. K. (2013). A Personalized recommendation-based mobile learning approach to improving the reading performance of EFL students. *Computers & Education*, *63*, 327-336.

Hussain, M., Zhu, W., Zhang, W., Abidi, S. M. R., & Ali, S. (2019). Using machine learning to predict student difficulties from learning session data. *Artificial Intelligence Review*, *52*(1), 381-407.

Jensen, J. L., Holt, E. A., Sowards, J. B., Ogden, T. H., & West, R. E. (2018). Investigating strategies for pre-class content learning in a flipped classroom. *Journal of Science Education and Technology*, *27*(6), 523-535.

Jovanović, J., Gašević, D., Dawson, S., Pardo, A., & Mirriahi, N. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *The Internet and Higher Education*, *33*(4), 74-85.

Kahu, E. R. (2013). Framing student engagement in higher education. *Studies in Higher Education*, *38*(5), 758-773.

Klašnja-Milićević, A., Vesin, B., Ivanović, M., & Budimac, Z. (2011). E-Learning personalization based on hybrid recommendation strategy and learning style identification. *Computers & Education*, *56*(3), 885-899.

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

Louhab, F. E., Bahnasse, A., Bensalah, F., Khiat, A., Khiat, Y., & Talea, M. (2020). Novel approach for adaptive flipped classroom based on learning management system. *Education and Information Technologies, 25*(2), 755-773. doi:10.1007/s10639-019-09994-0

Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender system application developments: A Survey. *Decision Support Systems*, *74*, 12-32.

McLaughlin, J. E., Griffin, L. M., Esserman, D. A., Davidson, C. A., Glatt, D. M., Roth, M. T., Gharkholonarehe, N., & Mumper, R. J. (2013). Pharmacy student engagement, performance, and perception in a flipped satellite classroom. *American Journal of Pharmaceutical Education*, *77*(9). doi:10.5688/ajpe779196

Montaner, M., López, B., & De La Rosa, J. L. (2003). A Taxonomy of recommender agents on the internet. *Artificial intelligence review*, *19*(4), 285-330.

Neto, J. L., Freitas, A. A., & Kaestner, C. A. (2002). Automatic text summarization using a machine learning approach. In *Brazilian symposium on artificial intelligence* (pp. 205-215). Berlin, Heidelberg: Springer. doi:10.1007/3-540-36127-8_20

Obeid, C., Lahoud, I., El Khoury, H., & Champin, P. A. (2018). Ontology-based recommender system in higher education. In *Companion Proceedings of 2018 Web Conference Companion (WWW 2018)* (*pp.* 1031-1034). Lyon, France: ACM. doi:10.1145/3184558.3191533

Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). E-Book-based learning analytics in university education. In *Proceeding of 23rd International Conference on Computer in Education (ICCE 2015)* (pp. 401-406). Hangzhou, China: APSCE.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vandeplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *The Journal of machine Learning research*, *12*, 2825-2830.

Persky, A. M., & Hogg, A. (2017). Influence of reading material characteristics on study time for pre-class quizzes in a flipped classroom. *American journal of pharmaceutical education*, *81*(6), 103. doi:10.5688/ajpe816103

Sams, A. (2011). The Flipped class: Shedding light on the confusion, critique, and hype. *The Daily Rift*, 1.

Shimada, A., Okubo, F., Yin, C., & Ogata, H. (2017). Automatic summarization of lecture slides for enhanced student previewtechnical report and user study. *IEEE Transactions on Learning Technologies*, *11*(2), 165-178.

Sun, J. C. Y., Wu, Y. T., & Lee, W. I. (2017). The Effect of the flipped classroom approach to OpenCourseWare instruction on students' self-regulation. *British Journal of Educational Technology*, *48*(3), 713-729.

Tarus, J. K., Niu, Z., & Mustafa, G. (2018). Knowledge-based recommendation: A Review of ontology-based recommender systems for e-learning. *Artificial intelligence review*, *50*(1), 21-48.

Yang, C. C., Chen, I. Y., Huang, A. Y., Lin, Q. R., & Ogata, H. (2020). Can self-regulated learning intervention improve student reading performance in flipped classrooms? *International Journal of Online Pedagogy and Course Design (IJOPCD)*, *10*(4), 1-13.

Yang, C.C., Akçapinar, G., Flanagan, B, & Ogata, H. (2019). Developing e-book page ranking model for pre-class reading recommendation. In *Proceeding of 27th International Conference on Computer in Education (ICCE 2019)* (pp. 360-362). Kenting, Taiwan: APSCE.

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.

The Relationships among Students' Personal Innovativeness, Compatibility, and Learning Performance: A Social Cognitive Theory Perspective

Wei-Tsong Wang* and Ying-Lien Lin
Department of Industrial and Information Management, National Cheng Kung University, Taiwan // wtwang@mail.ncku.edu.tw // r38021019@gs.ncku.edu.tw
*Corresponding author

(Submitted July 31, 2020; Revised August 21, 2020; Accepted January 12, 2021)

ABSTRACT: Although online learning systems (OLSs) are widely discussed with regard to most forms of higher education, they are still immature in terms of their development and compared to mainstream teaching and learning activities, and their implementation remains a challenge. Few studies have investigated the factors that contribute to students' learning performance in the context of the adoption of OLSs in higher educational institutions. Additionally, among the handful of studies that focus on this research issue, investigating the influences of environmental factors and personal innovativeness on students' learning performance and adoption of OLSs has received very little attention. Consequently, by integrating social cognitive theory and innovation diffusion theory, this study developed a theoretical model of OLS adoption. It empirically validated the model using data collected from 151 undergraduate students who used OLSs. The results showed that key compatibility (as an environmental factor) and personal innovativeness (as a personal factor) had significant direct and/or indirect influences on students' learning performance and continued intentions to use OLSs. Theoretical and practical implications are also discussed.

Keywords: Social cognitive theory, Innovation diffusion theory, Personal innovativeness, Compatibility, Learning performance

1. Introduction

The use of information technology systems (ITSs) via electronic devices can help students accomplish learning tasks. ITSs can transform how students learn new things and enable them to obtain the sophisticated domain knowledge that is needed to develop the higher-order critical thinking skills and necessary to cope with various conditions in real-world practices. Such skills are not easy to obtain. Using online learning systems (OLSs) may help students in higher educational institutions learn more effectively by providing useful functions to complement traditional, classroom-based educational arrangements. OLSs have various advantages for students, including accelerating the delivery of materials, accessing course content, increasing the flexibility and convenience of schedules, and removing geographic constraints (e.g., Hsu, Chen, & Ting, 2018; Novotny, Stapleton, & Hardy, 2016). Many universities have been using OLSs (i.e., Moodle, Blackboard) to facilitate students' learning and deliver online courses to improve students' learning performance or processes (e.g., Cheng & Yuen, 2019; Damnjanovic, Jednak, & Mijatovic, 2015; Risner & Kumar, 2016).

In the higher education field, a substantial body of studies has investigated OLSs. The purpose of this study is to investigate students' learning performance and continued intention in online learning based on the model of social cognitive theory (SCT). Few studies aim to understand the relationships among personal factors, environmental factors, and behavioral factors in the context of online learning. Therefore, the proposed model of this study includes two paths, namely, the person-belief-behavior path (personal innovativeness → perceived usefulness → continued intention or learning performance) and the environment-belief-behavior path (compatibility → perceived usefulness → continued intention or learning performance). Some researchers have examined the effects of compatibility and innovativeness on perceived usefulness and continued intention (e.g., Cheng, 2014; Huang, Yu, Tang, & Chang, 2019; Wang, Jung, Kang, & Chung, 2014). However, most of them do not address students' OLS learning experiences in ways that can improve students' learning performance in healthcare education settings. We extend the extant research and investigate personal innovativeness and compatibility, affecting students' learning performance and continued intention.

OLSs allow for active and self-directed learning as they provide students with the freedom to conveniently access, share, and discuss subject information, lecture notes, and other learning resources (Smart, Ross, Carollo, & Williams-Gilbert, 2020). OLSs have various advantages for students over conventional methods, including accelerating course content delivery, increasing the flexibility and convenience of schedules, removing geographic constraints, and promoting students' learning performance and continued intention (Acosta et al., 2018; Huang et al., 2019). Nevertheless, the adoption of OLSs among students has not yet reached the expected level. An OLS with a high level of compatibility means that its functions or innovation can meet students' preferences or needs, making them more comfortable for students to use. Moreover, each online course has unique course objectives or demands that may require unique solutions through OLSs (Smith, Passmore, & Faught, 2009). Such solutions facilitate students' willingness to continue using OLSs (Panigrahi, Srivastava, & Sharma, 2018). It is critical to identify the environmental factors (e.g., compatibility and perceived usefulness) of OLSs and students' innovativeness in higher educational contexts.

There have not been enough research efforts to investigate the personal or environmental factors of learning technologies that can better meet students' needs in higher educational contexts, especially in the areas of healthcare and nurse education (e.g., Smart et al., 2020; Zayim & Ozel, 2015). Prior studies have often been conducted from a technology-acceptance or motivation-theoretical perspective, and their contributions have thus been limited in providing a comprehensive understanding of the current study's focal issue. Therefore, this study adopts social cognitive theory (SCT), a comprehensive theory for understanding various types of human behaviors, as its primary theoretical base. We incorporate innovation diffusion theory (IDT) (Rogers, 1995; Rogers, 2003) to identify key personal factors and critical environmental factors based on the central premise of SCT to develop our research model. Additionally, the findings of this study can offer further insight for the developers of learning systems into the development of guidelines that enable the design and delivery of high-quality OLSs, including effective online mechanisms for students to obtain better learning outcomes. The results of the current can also assist policymakers in effectively allocating educational resources and enable instructors to teach in a manner that better supports students' learning activities. The research question (RQ) of this study is thus summarized as follows:
RQ: How do student-related personal, environmental, and behavioral factors influence students' learning performance in online courses supported by OLSs?

2. Review of the relevant literature

2.1. OLSs in higher education

Academics use various terms, including e-learning, distance learning, m-learning, virtual classroom, virtual learning environment, and web-based learning, which are often used interchangeably to denote online learning systems/tools (e.g., Panigrahi et al., 2018; Raman, Achuthan, Nedungadi, Diwakar, & Bose, 2014). These technology-supported learning systems are also called learning management systems, e-learning systems, or OLSs (Hill, Chidambaram, & Summers, 2017). In this study, we use the term "online learning system (OLS)."

The online courses supported by OLSs offer learner-centered instruction that can facilitate students' active learning, collaboration, and communication based on individual levels of innovativeness. Such OLS courses, especially clinical training courses, enhance students' self-confidence to improve their professional cognitive reasoning abilities to make decisions or judgments when they experience a significant challenge in real-world problem-solving. However, challenges remain for the effective use of OLSs. OLS developers must address challenges regarding all forms of special learning needs, such as the design of user-friendly interfaces and the sequence of learning tasks, for the sake of students. Additionally, design issues related to various platforms (e.g., iOS and Android); hardware manufacturers (e.g., Apple, Google, and Samsung); unfriendly user interfaces; and usability limitations must be carefully considered (Song, Singleton, Hill, & Koh, 2004). In online courses, students constantly utilize visual imaging tools and respond to simulated scenarios to enhance their cognitive reasoning abilities. Thus, the developers of OLSs should develop useful and compatible systems to assist students' learning tasks. Compatibility refers to students' fitness of beliefs, learning preferences, and values (e.g., critical thinking, information, and communication skills) with regard to online courses using OLSs (Cheng, 2014).

Al-Azawei, Parslow, and Lundqvist (2017) stated that eliminating learning barriers, matching learners' needs, and integrating instructional technology into e-course designs can improve learners' perceptions and beliefs. Consequently, Hansen (2018) argued that online learning with a flexible and self-paced model is an excellent way of accommodating students' learning needs from the point of view of the online competency-based education model. To address students' concerns regarding the use of an OLS, the OLS must provide a positive experience for its users and a positive linkage between students' learning performance and their willingness to continuously use OLSs.

There are practical difficulties for students in using OLSs, such as environmental factors and personal factors, which create significant challenges to instructors and OLS developers (Ma & Lee, 2019; Song et al., 2004). Several researchers have reported that the primary barriers to OLSs are administrative issues, technological issues, social influences, and individual differences rather than the design of the learning materials themselves (e.g., Ali, Uppal, & Gulliver, 2018; Mckimm, Jollie, & Cantillon, 2003). Such barriers may decrease the level of students' acceptance of OLSs. Inappropriate technologies, inadequate facilities (e.g., file format compatibility, video size/clip, and storage), insufficient technical support, a lack of learning skills, and a lack of training programs are the primary barriers to OLS success (Al-Azawei et al., 2017).

The adoption of OLSs by students is becoming increasingly popular because OLS-supported online learning tools and activities allow students to gain knowledge beyond the textbook and provide profound and meaningful learning experiences. From a nontechnological perspective, several studies have examined the effects of beliefs and characteristics on students' continued intentions (Cheng, 2014; Damnjanovic et al., 2015). Consequently, we believe that environmental factors and personal innovativeness are key factors influencing students' learning performance in OLS use.

2.2. The integrated theoretical model for building positive learning behaviors

The central premise of SCT is that there is a relationship between cognitive or personal factors, environmental factors, and behavior. Currently, SCT is widely used as a theoretical model in various fields such as academic achievement (Hill et al., 2017) and with regard to learners' continued intention to use an OLS (Iqbal & Bhatti, 2017). Based on the above discussion, we believe that SCT is an adequate framework for understanding students' learning performance because it incorporates the effects of both personal and environmental factors on the continued intention to use OLSs.

Additionally, it is argued that IDT, which is grounded in the sociological perspective, is a useful model for understanding the process by which the use of innovations spreads within and between social systems (Rogers, 1995). Agarwal and Prasad (1998) pointed out that personal innovativeness has diverse forms of effects on adopting new technologies. Personal innovativeness can be integrated into research models to measure educational innovation, which is an efficient way to analyze users' critical factors. Moreover, OLSs with a high compatibility level that matches users' needs can lead to high user-perceived usefulness levels (Cheng, 2014).

Because every theory has distinct roots and, thus, limited explanatory power, a theoretical framework that appropriately integrates various theoretical perspectives to comprehend a specific research issue is beneficial in advancing our understanding of the focal issue. Based on the discussion presented above, this study adopts SCT as the theoretical foundation and integrates IDT as part of the theoretical framework to extend the SCT model to investigate students' learning performance and behavioral intentions regarding OLS use in the context of higher education.

3. Research model and hypotheses

Based on previous studies, the SCT framework assumes relationships with its three factors associated with students adopting OLSs and learning performance in an online learning environment (e.g., Ifinedo, 2017; Iqbal & Bhatti, 2017). Although SCT is a triadic reciprocity model, we only used three SCT factors to examine students' continued intention and learning performance. Thus, the proposed hypotheses have no path loops. This study is a cross-sectional investigation and mainly examines the effects of two paths, i.e., the person-belief-behavior path and the environment-belief-behavior path, as shown in Figure 1.

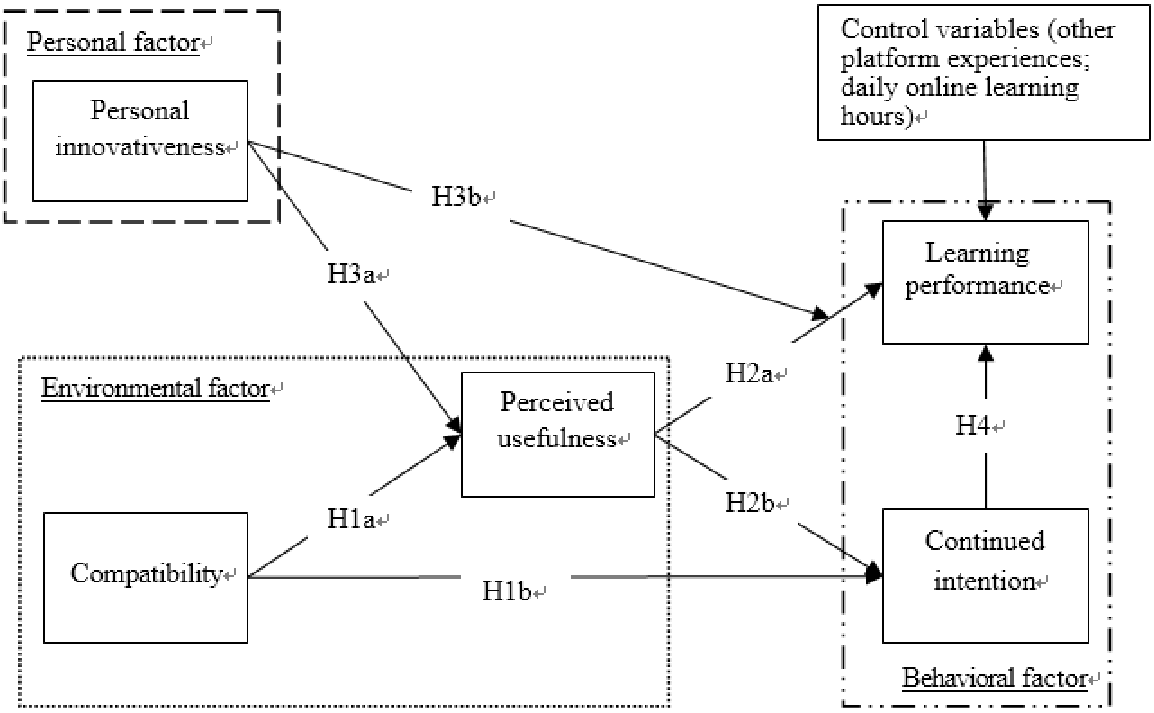


Figure 1. Proposed integrated model for online learning system adoption

3.1. Environmental factor

Based on SCT, Rogers (2003) defined compatibility as the degree to which students' use of OLSs is compatible with their current values, prior experiences, and current needs. *Perceived usefulness* can be described as the degree to which students perceive the adoption of an OLS as useful and appropriate in supporting their learning process (Park, 2009).

Compatibility is commonly regarded as a necessity for students to prevent potential conflict due to the different characteristics of various computers and learning systems (e.g., capabilities and applications) in their learning process. When an OLS has a high level of compatibility with students' preference and practice style, the willingness of those students to use OLSs will increase. This implies that compatibility can significantly affect perceived usefulness. The positive perceived value regarding the compatibility of an OLS is consistent with students' learning experiences and needs, which implies that an OLS with a high level of compatibility leads to increased continued intention (Raman et al., 2014; Cheng, 2014). Accordingly, the following hypotheses are proposed:
H1a Compatibility is positively related to the perceived usefulness of an OLS.
H1b Compatibility is positively related to continued intention to use an OLS.

The research indicates that perceived usefulness significantly affects students' learning performance (Davis, 1989). A useful OLS can provide a better learning environment and assist students in managing their learning process. The perceived usefulness of an OLS means that students' perception that OLSs will enhance their learning performance (Johnson, Hornik, & Salas, 2008; Liaw & Huang, 2016). It can be inferred that perceived usefulness is a key determinant of students' judgments regarding their learning performance in an online course as a high level of perceived usefulness will enhance students' beliefs of outcome values. Additionally, the relationship between students' perceived usefulness and their continued intentions to use an OLS has been extensively studied (e.g., Findik-Coşkunçay, Alkiş, & Özkan-Yildirim, 2018; Huang et al., 2019; Ma & Lee, 2019). Thus, the essence of perceived usefulness (e.g., a well-designed system, perceived value, and service quality) can positively affect students' continued intention to use an OLS. Therefore, the following hypotheses are proposed:
H2a Perceived usefulness is positively related to students' learning performance.
H2b Perceived usefulness is positively related to continued intention to use an OLS.

3.2. Personal factor

Individuals with high levels of personal innovativeness may be risk-takers, and, therefore, they may focus their attention on the performance of new technology (Liu, Li, & Carlsson, 2010). Innovative students are more likely to develop more positive beliefs about OLSs than those with lower levels of innovativeness. Students focus on their preferred learning pathways, which they perceive to enable them to gain various experiences and change their learning behaviors. Such positive beliefs regarding OLS performance are likely to increase students' perceptions of OLS usefulness (e.g., Huang et al., 2019; Lu, Yao, & Yu, 2005). Based on the results of these studies, this study considered personal innovativeness to be a determinant of perceived OLS usefulness. Therefore, the following hypothesis is proposed:
H3a Personal innovativeness is positively related to the perceived usefulness of OLSs.

In the current study, innovativeness refers to an individual's inherent tendency to try new technologies (Agarwal & Prasad, 1998). Individuals with a high level of innovativeness tend to have positive performance expectancy because they are more likely to utilize new learning technologies well, to identify and access multiple sources of critical knowledge, to adopt new ideas, and then change learning styles or habits correspondingly and to adjust self-learning strategies to achieve better performance. The prior studies show that personal innovativeness is a key moderator of the effect of perceived usefulness on IT usage behavior and its consequences (e.g., Agarwal & Prasad, 1998; Cheng, 2014; Shaw & Sergueeva, 2019). Therefore, the following hypothesis is developed:
H3b Personal innovativeness positively moderates the relationship between perceived usefulness and learning performance.

3.3. Behavioral factor

In this study, continued intention refers to students' intention to use an IT-based tool (e.g., an OLS) to perform learning activities (Davis, 1989). When individuals have a pleasant experience using OLSs, they may exhibit a positive attitude toward online learning activities and are more likely to continue to use OLSs to assist them in learning. If the functions provided by OLSs are a good match with students' learning tasks, the use of those OLSs may improve their learning performance. Students will likely use an OLS to achieve a desirable learning performance (Han &Yi, 2019; Kuo & Hwang, 2015). Therefore, the following hypothesis is proposed:
H4 Students' continued intention to use OLSs is positively related to learning performance.

3.4. Control variables

Daily internet hours and experience with other online learning platforms were used as the control variables for the constructs of learning performance to avoid spurious effects and specification errors in the model as they are contextual variables that have the potential to influence issues related to ITSs.

4. Methodology

4.1. Participants and procedure

In this study, a survey approach was used to collect data about the experiences of students who enrolled in online courses. The courses used in this study are online curricula because most of the lectures are delivered online using OLSs. The lectures are organized into weekly lessons and alternate between online (14 weeks) and face-to-face (4 weeks) formats. The instructional delivery approach of the online courses is designed following school regulations. Although the online courses are based on blended curricula (e.g., Jonker, März, & Voogt, 2020; Risner & Kumar, 2016), they emphasize the online lecture delivery modality (Smith et al., 2009).

The vast majority of studies assessing online learning adoption tacitly postulate that adoption among disciplines of online learning is homogeneous (Smith et al., 2009). In this study, the online courses investigated are part of university curricula and cover essential healthcare subjects. The qualifications of the students enrolled in online courses are limited to a specific semester or department. Additionally, the online courses use similar formats of teaching materials, processes, and digital tools as those supported by OLSs (e.g., Novotny et al., 2016; Teo, Zhou, Fan, & Huang, 2019; Cheng & Yuen, 2019). The functions of OLSs include text-based content, multimedia, prerecorded video, hyperlinks, face-to-face lectures, traditional assessment methods (e.g., exams), and communication tools (e.g., messages, e-mail, forums). We thus consider the online courses investigated in this study to be essentially homogeneous.

The survey participants were students who enrolled in six courses related to optometry, biostatistics, healthcare, medical computer applications, and multimedia design for geo-information (two classes) in two medical universities. Unlike the conventional instructional delivery approach, these online courses must be approved after a school meeting. The unique contexts require an understanding that online courses may be similar but that each student selects online courses with a slightly different set of goals and objectives. An online course can provide flexibility for students to learn at a convenient time and enhance their learning effectiveness in conjunction with other learning strategies (Smart et al., 2020). For example, students may have the opportunity to participate in an off-campus internship program in hospitals or different industries. Additionally, students can cover more material in a shorter amount of time in online courses. Further, online courses are convenient for those who must pass a national examination for licensure as an optometrist, pharmacist, nurse, or other healthcare professional. Thus, the content of online courses can be delivered more effectively to students than through conventional learning methods.

According to our investigation, approximately 50 to 60 classes are organized (approximately less than 2% of the three thousand courses every semester) as asynchronous or synchronous online courses at one of the universities but only 10 or fewer courses at the other university every semester. An invitation letter was sent to the instructors of all 10 online courses offered to request their participation. The instructors of six online courses (with approximately 300 potential participants) agreed to participate in the study. All of the students in those six courses were invited to participate in our survey, and the course instructors agreed to provide grades from the midterm and final exams for our data analysis procedures with the consent of the participating students. The data were collected using an online survey by Google Forms, which was analyzed to validate the proposed model shown in Figure 1. We used the LINE app or an email, and the questionnaires were then distributed to the students of six courses. In addition, the instructors were informed of the nature of the survey. We did not force the students to participate in the survey. If the students were not willing to provide their scores, they were not required to click the URL of the Google Form. The participants were asked to answer the survey questionnaire based on their experience in the online learning process.

The qualifications of the participants were ensured, and the instructors provided the respondents' midterm and final exam scores based on the respondents' IDs. All research procedures were conducted with the consent of the instructors who had reviewed the questionnaire beforehand. There was a short description of the study at the beginning of the questionnaire so the participants could fully understand their rights and interests. Thus, all the participants were aware that their participation was voluntary and that they could freely participate or withdraw at any time. All the respondents remained anonymous in the survey. Ethics approval of the study's procedures (Ethics Committee No. 109-088-02) was obtained through the authors' university governance framework for human research.

4.2. Instrument development

With regard to the content of the questionnaire, we invited two scholars in the field of online learning and education to examine the questionnaire. The wording of the question items was slightly modified to fit the online learning context and enhance validity. Personal innovativeness was evaluated by using the three items from Agarwal and Prasad (1998). Compatibility was measured by using the three-item scale of Duan, He, Feng, Li, and Fu (2010). Perceived usefulness was assessed by using the three-item scale of Davis et al. (1989). Continued intention was measured by using the three-item scale of Cheng (2014). The learning performance construct was evaluated based on the scores of the students' midterm and final exam, which were provided by the courses' teachers. Except for learning performance (students' midterm exam and final exam scores range from 1 to 100), other items were rated using a five-point Likert scale in which 1 means "strongly disagree" and 5 means "strongly agree."

4.3. Data analysis procedures

This study mainly concerns the effects of personal innovativeness and compatibility on learning performance. In the survey, the covariance-based structural equation modeling (CB-SEM) method was used to evaluate the structural relationships among the variables. Byrne (2006) suggested that the CB-SEM approach is appropriate and exhibits robust statistics and fit indices against the normality assumption for theory-driven empirical research. Thus, the PROCESS macro of SPSS software developed by Hayes (2013) was used to perform the data analysis procedures for evaluating the moderating effect of personal innovativeness. Additionally, the AMOS software package was used to conduct the CB-SEM technique's data analysis procedures to examine the goodness-of-fit of the measurement and structural models of the research model and validate the developed hypotheses.

Reflective measurement instruments were used to evaluate reliability, convergent validity, and discriminant validity to ensure that their corresponding constructs were adequate. Composite reliability (CR) was used to assess scale reliability, while Cronbach's alpha coefficients were used to evaluate internal consistency reliability. The reliability of the threshold was over .7 for CR; a value of .7 for Cronbach's alpha is recommended (Hair, Black, Babin, & Anderson, 2010). If the factor loadings of the indicators are beyond the restrictive criterion of .5 or higher, they are significantly related to the corresponding constructs. If the average variance extracted (AVE) values of the reflective constructs are higher than .5, convergent validity is supported. If the squared correlations between a given construct and other constructs in the model are smaller than the corresponding AVE estimates, it suggests that the measurements are adequate for research purposes. In addition, the values of the square root of the AVE are greater than the correlation estimate of the two constructs, which shows good evidence of discriminant validity.

Finally, an adequate structural model was obtained, and the significance of our research hypothesis was obtained using CB-SEM to estimate the parameter via 151 valid samples. The indices for examining the goodness-of-fit of both the measurement and structural models of our research model were examined using those recommended for CB-SEM research, including the product of the minimum-fit-function chi-square statistic divided by the degree of freedom (χ^2/df), robust confirmatory fit index (CFI), root-mean-square error of approximation (RMSEA), standardized root mean square residual (SRMR), goodness-of-fit index (GFI), adjusted goodness-of-fit index (AGFI), normed fit index (NFI), and Tucker–Lewis index (TLI). Additionally, we examined the significance of the path coefficient estimation and the determination coefficients (R^2) of the endogenous constructs.

5. Data analysis and results

A total of 18 responses were considered invalid due to the respondents' failure to properly answer the survey questions or to obvious systemic answers, thus yielding 151 valid responses. Regarding our sample's demographic profile, 60.26% of the respondents were female; this profile is similar to those of the students at the two universities that participated in this study. All of them were between 19 and 23 years old, and more than 86.09% of them spent more than two hours per day in online learning. Additionally, 27.81% of the students had experience with other learning platforms. *For the control variables*, the results showed that daily internet/online learning hours were significantly associated with learning performance, whereas experience with other online learning platforms was not.

Table 1. The results of goodness-of-fit measure via CB-SEM

Model	χ^2	χ^2/d.f.	RMSEA	SRMR	CFI	GFI	AGFI	NFI	TLI
Measurement	107.45	1.6	.06	.04	.97	.92	.87	.93	.96
Structural	169.29	2.42	.1	.07	.94	.87	.81	.9	.92
Cutoff value	-	< 3	< .08-.1[(a)]	< .08	> .9	> .9	> .8	> .9	> .9

Note. [(a)] The criteria are acceptable (Hu & Bentler, 1999).

Measurement model. First, the content validity was established before we used the confirmatory factor analysis (CFA) technique. All constructs of the proposed model were first-order reflective constructs, and all items were based on the empirical theories of previous studies (e.g., Agarwal & Prasad, 1998; Davis et al., 1989; Duan et al., 2010; Hill et al., 2017; Ifinedo, 2017). Second, to examine the level of goodness-of-fit (GoF) of our measurement model, we examined multiple GoF indices, as presented in Table 1, and the results indicated an adequate GoF level for the proposed measurement model according to the recommended threshold values (Hair et al., 2010). Finally, the construct validity was evaluated based on the following criteria (Hair et al., 2010). As shown in Table 2, all factor loadings, ranging from .65 to .96, were statistically significant (p < .000) and were all greater than the cutoff value of .5. Additionally, the AVE estimates of all constructs, ranging from .62 to .85, were greater than the cutoff value of .5. Furthermore, as presented in Table 3, we found that all the square roots of AVE estimates for all pairs of constructs were higher than the correlation between the two constructs. Moreover, as presented in Table 3, the CR statistics for all constructs, ranging from .77 to .94, were higher than the cutoff value.7. Finally, Cronbach's alpha statistics of all the constructs ranged from .75 to .94, which were all greater than the cutoff value.7.

Table 2. Results of the measurement model

Construct	Items	Factor loading
Personal innovativeness (PINN)	If I heard about a new online course, I would look for a way to gain experience with it.	.65
	Among my peers, I am usually the first to try out a new online course.	.91
	I like to experiment with new online courses.	.89
Compatibility	Using OLSs provided by the school for my learning is more suitable for my learning style.	.77
	Using OLSs provided by the school is more suitable for my lifestyle.	.95
	The learning format of OLSs provided by the school meets my learning needs very well.	.85
Perceived usefulness (PU)	Using online courses improves my learning performance.	.78
	Using online courses enhances my learning efficiency.	.83
	Using online courses can help me increase my learning effectiveness.	.83
Continued intention	I intend to use online courses to perform my learning activities and communicate with my classmates.	.92
	I would use online courses to perform different learning-related activities.	.96
	I intend to increase my use of online courses in the future.	.88
Learning performance	Your midterm exam score is:	.72
	Your final exam score is:	.85

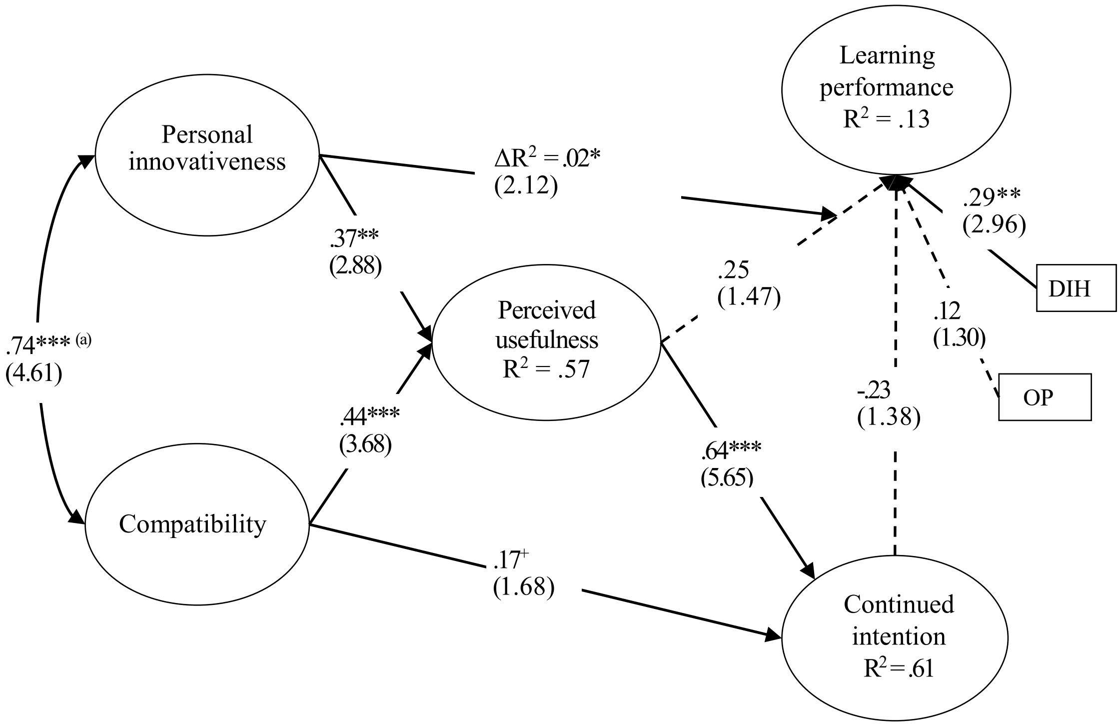
Note. All factor loadings are significant at $p < .001$ according to CB-SEM.

Table 3. Descriptive statistics, correlations, AVE, CR, and Cronbach's alpha

	1	2	3	4	5
1. Compatibility	**.86**				
2. Personal innovativeness	.65**	**.88**			
3. Perceived usefulness	.56**	.57**	**.81**		
4. Continued intention	.5**	.51**	.69**	**.92**	
5. Learning performance	.06	-.06	.09	-.05	**.79**
Mean	3.77	3.72	3.86	3.7	73.63
SD	.63	.66	.66	.74	15.93
Cronbach's alpha	.89	.85	.86	.94	.75
Composite reliability (CR)	.89	.91	.85	.94	.77

Note. $N = 151$, *SD* standard deviation. The square root of AVE is on the diagonal and the other matrix entries are the factor correlations.

Structural model. We first examined the level of the GoF of our structural model by examining multiple GoF indices, as presented in Table 1 above, and the results indicated an adequate level of goodness-of-fit of our structural model according to the recommended threshold values (Hair et al., 2010). Additionally, the criterion validity of the structural model was ensured in two ways. We used the predictive relevance of the correlation between a measure developed based on a reflective perspective and the external criterion (i.e., dependent variable). R^2 was used to predict the accuracy of our structural model. The three levels of R^2 for endogenous constructs can be measured as follows: .75 for substantial, .5 for moderate, and .25 for weak (Hair, Ringle, & Sarstedt, 2011). The results show that the R^2 values of our three endogenous constructs, namely, perceived usefulness, learning performance, and continued intention, were equal to .57, .13, and .61, respectively. The findings indicate that our research model exhibited weak to moderate predictive accuracy. These results reveal that our proposed model had medium to large predictive relevance for the three endogenous constructs. Overall, the results of this study based on the examinations above support the criterion validity of the research model. Figure 2 shows that compatibility is positively and significantly related to perceived usefulness (H1a: $\beta = .44$, $p < .001$) and continued intention (H1b: $\beta = .17$, $p < .01$). Perceived usefulness is positively and significantly related to learning performance (H2a: $\beta = .25$, $p > .1$) and continued intention (H2b: $\beta = .64$, $p < .001$). Personal innovativeness is positively and significantly related to perceived usefulness (H3a: $\beta = .37$, $p < .01$) and has a positive and significant moderating effect on the relationship between perceived usefulness and learning performance (H3b: R^2 *changed* = .02, $t = 2.12$, $p < .05$). Additionally, the bias-corrected 95% confidence interval with 5000 bootstrapping tests confirmed that the moderation coefficient exactly excluded zero (.36-10.11), representing a significant moderating effect of personal innovativeness. Moreover, Figure 3 shows the moderating effect's slope, which indicates that the level of personal innovativeness significantly increases the positive influence of perceived usefulness on learning performance. Continued intention is negatively but insignificantly related to learning performance (H4: β = -.23, p > .1). Therefore, all hypotheses are supported, except for H2a and H4.



***p< .001; **p< .01; *p< .05; +p< .1; T-values are in parentheses.
DIH daily internet/online learning hours, *OP* experience with other online learning platforms.
(a) Correlation parameter.

Figure 2. Hypotheses testing results of the proposed research model

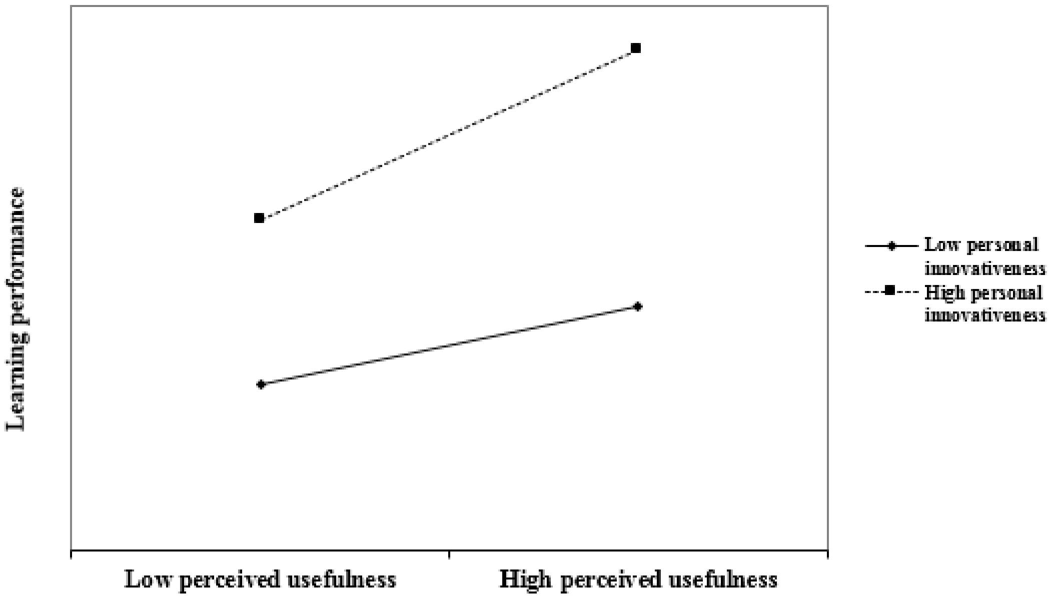


Figure 3. The moderating effect of personal innovativeness on the relationship between perceived usefulness and learning performance

6. Discussion

The findings of this study have confirmed that IDT is incorporated into the SCT model, which produces valuable insights as a framework to use for online healthcare courses. Personal innovativeness and compatibility enhance student beliefs by using OLSs and thus increase students' learning performance. The first path (person-belief-behavior) is supported by personal innovativeness → perceived usefulness → learning performance/continued intention. Simultaneously, the second path (environment-belief-behavior) is supported by compatibility → perceived usefulness → learning performance/continued intention. SCT is an influential theory that can help educators/instructors/students efficiently plan online teaching/learning modules.

With regard to the environment-belief-behavior path, our findings show that compatibility is positively related to perceived usefulness (H1a) and continued intention (H1b), which is consistent with the findings of previous studies (e.g., Cheng, 2014; Raman et al., 2014). The result means that the usefulness of an OLS is determined by its ability to support students' self-directed learning, preferences, and peer interactions. This study suggests that students generally have a high level of compatibility in online educational settings, which results in higher loyalty to online learning and a positive attitude toward online courses.

Additionally, perceived usefulness is positively yet insignificantly related to learning performance (H2a), which is not consistent with the findings of previous studies (e.g., Davis, 1989; Findik-Coşkunçay et al., 2018; Han & Yi, 2019; Liaw & Huang, 2016). Thus, we used independent *t*-test analysis to compare students' perceptions of OLS usefulness based on groups with high learning performance and those with low learning performance. The results revealed an insignificant difference between both groups; however, a tendency was observed. It is thus reasonable to infer that groups of students with low learning performance certainly depend on OLSs to support their learning while those with high learning performance have high capabilities to cope with online courses and, thus, may not be impressed when using OLSs. Consequently, perceived usefulness is positively and significantly related to continued intention to use OLSs (H2b), which is in line with the findings of previous studies (e.g., Teo et al., 2019; Wang et al., 2014). The results show the significant role played by perceived usefulness. This finding provides novel insights to promote an OLS to improve the learning of healthcare students with implications for instructors and students in healthcare education.

With regard to the person-belief-behavior path, the result indicates that students' innovativeness is positively related to their perceived usefulness (H3a), which is in line with the findings of previous studies (Huang et al., 2019; Liu et al., 2010). Additionally, personal innovativeness has a significant moderating effect that increases the positive relationship between perceived usefulness and learning performance (H3b). This finding is first confirmed in this study. According to the personal factors of SCT, innovative students often have relevant knowledge and the ability to use OLSs to learn. Such a situation may motivate them to adopt a self-directed learning method according to their learning pace or to utilize other website learning resources to improve learning performance by supplementing online courses. This study suggests that instructors should foster students' innovativeness so they will be more willing to adopt new learning technologies and further strengthen their confidence to improve their learning performance.

Moreover, continued intention to use OLSs is insignificantly related to learning performance (H4), which is inconsistent with the findings of previous studies. There are possible explanations for this result. Online learning often emphasizes self-directed learning; thus, students must frequently interact and communicate with their peers/instructors and obtain accurate knowledge or information content compared to face-to-face courses. Simultaneously, instructors must provide timely feedback in response to students' questions, help students fill professional knowledge gaps, and encourage them to develop a more in-depth understanding of course content. In fact, in healthcare education, some issues have not been fully addressed, such as the flexibility of OLSs, students' adaptability, and their proficiency with OLSs (Smart et al., 2020). Based on the personal, environmental, and behavioral factors of SCT, we suggest that future research provide more insights into the external factors of learning performance, which may enrich the research on OLS adoption.

Figure 2 shows that the control variables, such as the students' other platform experiences and the number of hours students spend on online learning each day, are insignificantly and significantly related to learning performance, respectively. This implies that students' experiences of other platforms do not provide unique benefits to support their learning interaction or communications in healthcare courses.

6.1. Research and practical implications

The present findings have some implications for practice and future research. Few studies have specifically investigated the relationship between students' continued intentions and their learning performance in higher education settings. First, a useful OLS can enhance students' interaction with instructors, peers, and materials, promoting students' engagement and facilitating cognitive reasoning abilities and critical thinking skills. From the perspective of SCT, our findings support the proposition of social cognitive learning that is manifested by the relationships among the human, environmental, and behavioral factors proposed by the SCT. Therefore, instructors can actively offer some supporting teaching materials (e.g., case-based solutions) in online courses, encourage students to interact, participate, and discuss closely with peers, and give students appropriate guidance for learning critical knowledge that is important to healthcare professionals.

Considering a personal factor, with regard to the study's findings, it can be reasonably inferred concerning personal innovativeness that traditional learning tools and pedagogical approaches may need to be changed to meet students' needs to actively seek new learning strategies, knowledge, or ideas through interaction. In such a case, students possess the capability to cope with a high level of uncertainty and accept new technology in the real world.

In considering environmental factors, the findings of compatibility and perceived usefulness show that the diffusion of technological innovation and the adoption of new technology have a high level of explanatory power and a complementary feature that is supported in this study. Our findings contribute to an understanding the element of compatibility with respect to students' capability of obtaining knowledge when they interact with an OLS in the online learning process.

Our findings also offer important practical implications. First, OLS developers must provide students with useful functions for enhancing intersystem connectivity and interpersonal communication (e.g., visual image retrieval from online databases and instant messaging functions) to satisfy healthcare students' learning needs. OLS developers can integrate several modules to develop new features based on user feedback and the results of the constant monitoring and evaluation of OLS performance. For example, OLSs can provide students with hyperlinks to access various online repositories or platforms to acquire critical information and quality tools to improve healthcare students' proficiency in acquiring professional skills.

Second, OLSs enable students to learn effectively in online courses at their own learning pace and based on their preferences. Therefore, course designers can design courses based on the context of healthcare training to enrich students' understanding of critical healthcare-related knowledge. Curriculum designers can use a variety of methods to embed a series of healthcare issues for each curriculum (e.g., medical humanities), allowing students to establish closer interactions and deeper discussion with lecturers and learning partners in online courses.

Finally, instructors must integrate relevant course materials (e.g., clinical films) and resources (e.g., sensitive medical records) into their online teaching. For example, instructors can combine previous clinical experience and basic scientific and medical principles into teaching content to increase students' interest in learning and enhance their reflection, creativity, and skill development. Additionally, instructors can encourage students to be open-minded and interact and share their expertise, experiences, or opinions with peers. These learning processes can help students improve their cognitive learning skills (Huang et al., 2019; Smith et al., 2009).

6.2. Limitations and future research

The limitations of this study are as follows. Concerning the generalizability of the findings, a cross-sectional analysis was performed based on questionnaires collected from 151 healthcare students from two medical and pharmacy universities. Future research could extend the proposed model to various OLSs or issues by collecting data from clinical training areas and using different sampling procedures. The results show that personal innovativeness and compatibility are critical antecedents that have significant positive direct or indirect relationships with students' continued intention and learning performance. However, the effects of other personal factors, including learning style, self-directed ability, and technology experience, could be considered in future research projects to enrich our understanding of students' learning performance in the context of OLS use.

7. Conclusion

Based on an integrated view, the relationships among the critical factors of three significant conceptual dimensions of SCT, namely, personal, environmental, and behavioral factors, were confirmed in this study. The findings of the study send a critical message to OLS developers, learning designers, and instructors. From the environmental and behavioral perspective, attracting and retaining healthcare students is of global concern for higher education institutions in OLSs. Universities can offer various lectures (e.g., blended or purely online) to respond to all students' needs while referring to the valuable findings of this study for their decision-making in the allocation of resources or educational strategies. From the personal and behavioral perspective, the development of OLSs should not stop here. Learning designers should be concerned with the requirements of innovative students' perceived usefulness and the compatibility of an OLS, improving course design to provide more efficient learning methods and content. In online healthcare courses, three interrelated core concepts (i.e., material, interaction, and embodiment) are bound together as priorities. From a holistic view of OLSs, instructors should eliminate the myth of traditional lectures and rethink the three factors of SCT to adjust teaching strategies in their online courses.

Acknowledgement

We want to thank the editor and two anonymous reviewers for their comments and suggestions. We also gratefully acknowledge Professor Hong-Long Chen and all people who provided valuable insights and information to this paper. This research was funded by the Ministry of Science and Technology of Taiwan (grant number: MOST 107-2511-H-006-005-MY2).

References

Acosta, M. L., Sisley, A., Ross, J., Brailsford, I., Bhargava, A., Jacobs, R., & Anstice, N. (2018). Student acceptance of e-learning methods in the laboratory class in Optometry. *PloS One*, *13*(12), e0209004.

Agarwal, R., & Prasad, J. (1998). The Antecedents and consequents of user perceptions in information technology adoption. *Decision Support Systems*, *22*(1), 15-29.

Al-Azawei, A., Parslow, P., & Lundqvist, K. (2017). The Effect of universal design for learning (UDL) application on e-learning acceptance: A Structural equation model. *The International Review of Research in Open and Distributed Learning*, *18*(6), 1-21.

Ali, S., Uppal, M. A., & Gulliver, S. (2018). A Conceptual framework highlighting e-learning implementation barriers. *Information Technology & People*, *31*(1), 156-180.

Byrne, B. M. (2006). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. London, UK: Lawrence Erlbaum Associates.

Cheng, M., & Yuen, A. H. (2019). Cultural divides in acceptance and continuance of learning management system use: A Longitudinal study of teenagers. *Educational Technology Research and Development*, *67*(6), 1613-1637.

Cheng, Y. M. (2014). Exploring the intention to use mobile learning: The Moderating role of personal innovativeness. *Journal of Systems and Information Technology*, *16*(1), 40-61.

Damnjanovic, V., Jednak, S., & Mijatovic, I. (2015). Factors affecting the effectiveness and use of Moodle: Students' perception. *Interactive Learning Environments*, *23*(4), 496-514.

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

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

Duan, Y., He, Q., Feng, W., Li, D., & Fu, Z. (2010). A study on e-learning take-up intention from an innovation adoption perspective: A Case in China. *Computers & Education*, *55*(1), 237-246.

Findik-Coşkunçay, D., Alkiş, N., & Özkan-Yildirim, S. (2018). A Structural model for students' adoption of learning management systems: An Empirical investigation in the higher education context. *Educational Technology & Society*, *21*(2), 13-27.

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

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, *19*(2), 139-152.

Han, S., & Yi, Y. J. (2019). How does the smartphone usage of college students affect academic performance? *Journal of Computer Assisted Learning*, *35*(1), 13-22.

Hansen, S. E. (2018). Improving educational and financial effectiveness through innovation: A Case study of Southern New Hampshire University's College for America. *International Journal of Educational Development*, *58*, 149-158.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A Regression-based approach*. New York, NY: Guilford Press.

Hill, T., Chidambaram, L., & Summers, J. D. (2017). Playing "catch up" with blended learning: Performance impacts of augmenting classroom instruction with online learning. *Behaviour & Information Technology*, *36*(1), 54-62.

Hsu, J. Y., Chen, C. C., & Ting, P. F. (2018). Understanding MOOC continuance: An Empirical examination of social support theory. *Interactive Learning Environments*, 1-19.

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. T., Yu, C. L., Tang, T. W., & Chang, S. C. (2019). A Study of the use of mobile learning technology in Taiwan for language learning. *Innovations in Education and Teaching International*, 1-13.

Ifinedo, P. (2017). Examining students' intention to continue using blogs for learning: Perspectives from technology acceptance, motivational, and social-cognitive frameworks. *Computers in Human Behavior*, *72*, 189-199.

Iqbal, S., & Bhatti, Z. A. (2017). What drives m-learning? An Empirical investigation of university student perceptions in Pakistan. *Higher Education Research & Development*, *36*(4), 730-746.

Johnson, R. D., Hornik, S., & Salas, E. (2008). An Empirical examination of factors contributing to the creation of successful e-learning environments. *International Journal of Human-Computer Studies*, *66*(5), 356-369.

Jonker, H., März, V., & Voogt, J. (2020). Curriculum flexibility in a blended curriculum. *Australasian Journal of Educational Technology*, *36*(1), 68-84.

Kuo, F. R., & Hwang, G. J. (2015). A Structural equation model to analyse the antecedents to students' web-based problem-solving performance. *Australasian Journal of Educational Technology*, *31*(4), 400-420.

Liaw, S. S., & Huang, H. M. (2016). Investigating learner attitudes toward e-books as learning tools: Based on the activity theory approach. *Interactive Learning Environments*, *24*(3), 625-643.

Liu, Y., Li, H., & Carlsson, C. (2010). Factors driving the adoption of m-learning: An Empirical study. *Computers & Education*, *55*(3), 1211-1219.

Lu, J., Yao, J. E., & Yu, C. S. (2005). Personal innovativeness, social influences and adoption of wireless Internet services via mobile technology. *The Journal of Strategic Information Systems*, *14*(3), 245-268.

Ma, L., & Lee, C. S. (2019). Investigating the adoption of MOOCs: A Technology-user environment perspective. *Journal of Computer Assisted Learning*, *35*, 89-98.

McKimm, J., Jollie, C., & Cantillon, P. (2003). ABC of learning and teaching web based learning. *BMJ: British Medical Journal*, *326*, 870-873.

Novotny, N. L., Stapleton, S. J., & Hardy, E. C. (2016). Enhancing critical thinking in graduate nursing online asynchronous discussions. *Journal of Nursing Education*, *55*(9), 514-521.

Panigrahi, R., Srivastava, P. R., & Sharma, D. (2018). Online learning: Adoption, continuance, and learning outcome - A Review of literature. *International Journal of Information Management*, *43*, 1-14.

Park, S. Y. (2009). An Analysis of the technology acceptance model in understanding university students' behavioral intention to use e-learning. *Journal of Educational Technology & Society*, *12*(3), 150-162.

Raman, R., Achuthan, K., Nedungadi, P., Diwakar, S., & Bose, R. (2014). The VLAB OER experience: Modeling potential-adopter student acceptance. *IEEE Transactions on Education*, *57*(4), 235-241.

Risner, M., & Kumar, S. (2016). Graduate student perceptions of a globally networked course. *Journal of Applied Research in Higher Education*, *8*(3), 287-301.

Rogers, E. M. (1995). Lessons for guidelines from the diffusion of innovations. *Joint Commission Journal on Quality and Patient Safety*, *21*(7), 324-328.

Rogers, E. M. (2003). *Diffusion of innovations* (5th Ed.). New York, NY: Free Press.

Shaw, N., & Sergueeva, K. (2019). The Mon-monetary benefits of mobile commerce: Extending UTAUT2 with perceived value. *International Journal of Information Management*, *45*, 44-55.

Smart, D., Ross, K., Carollo, S., & Williams-Gilbert, W. (2020). Contextualizing instructional technology to the demands of nursing education. *CIN: Computers, Informatics*, *Nursing*, *38*(1), 18-27.

Smith, G. G., Passmore, D., & Faught, T. (2009). The Challenges of online nursing education. *The Internet and Higher Education*, *12*(2), 98-103.

Song, L., Singleton, E. S., Hill, J. R., & Koh, M. H. (2004). Improving online learning: Student perceptions of useful and challenging characteristics. *The Internet and Higher Education*, *7*(1), 59-70.

Teo, T., Zhou, M., Fan, A. C. W., & Huang, F. (2019). Factors that influence university students' intention to use Moodle: A Study in Macau. *Educational Technology Research and Development*, *67*(3), 749-766.

Wang, T., Jung, C. H., Kang, M. H., & Chung, Y. S. (2014). Exploring determinants of adoption intentions towards Enterprise 2.0 applications: An empirical study. *Behaviour & Information Technology*, *33*(10), 1048-1064.

Zayim, N., & Ozel, D. (2015). Factors affecting nursing students' readiness and perceptions toward the use of mobile technologies for learning. *CIN: Computers, Informatics, Nursing*, *33*(10), 456-464.

Eutsler, L., & Long, C. S. (2021). Preservice Teachers' Acceptance of Virtual Reality to Plan Science Instruction. *Educational Technology & Society*, *24* (2), 28–43.

Preservice Teachers' Acceptance of Virtual Reality to Plan Science Instruction

Lauren Eutsler* and Christopher S. Long
University of North Texas, TX, United States // lauren.eutsler@unt.edu // chris.long@unt.edu
*Corresponding author

(Submitted October 1, 2020; Revised October 23, 2020; Accepted January 22, 2021)

ABSTRACT
To improve understanding of preservice teacher acceptance and integration of virtual reality into science, this study examined individual concerns to integrate virtual reality into science instruction before and after a hands-on intervention with virtual reality. Framed by the Concerns-Based Adoption Model using a mixed-method design, preservice teachers were exposed to a 5-week intervention to integrate and expand on existing VR tours and construct a personalized VR tour. Pre and post analysis of the stages of concern questionnaire show four of five preservice teachers remained focused on their personal concerns (stage 2, unsure of VR teaching demands). The fifth advanced to stage 3, management, and was interested in learning ways to implement virtual reality in the classroom. Open-ended data (survey items, science journals, focus group) illuminated concerns about the technical aspects of VR, learning engagement/satisfaction, and generation of lesson plan ideas, which influenced preservice teachers' intention to use VR. For four of five preservice teachers, this experience increased their likelihood to use VR in the classroom, with adoption dependent on using VR with their students. Implications for teacher educators, educational researchers, administrators, and digital designers address the integration of VR, instructional planning, and usability considerations.

Keywords: Curriculum design, Pedagogy, Science, Technology acceptance, Virtual reality

1. Introduction

Higher education students use digital technology most frequently for personal and informal reasons, to communicate socially with one another (Echenique et al., 2015; Gasaymeh, 2018). A synthesis of a special issue on virtual reality (VR) in learning revealed vivid and enhanced ways that students can learn with "virtual and augmented reality modes;" however, the synthesis stated that this shift would require "extensive research on content and context design" (Lytras et al., 2016, p. 878). Ioannou and Ioannou (2020) identified academic gains and positive perceptions when middle-schoolers' used VR to go on a virtual tour of Archaic kingdoms, but more research is needed at the intersection of technology, design, and pedagogy when integrating VR.

Other integration factors such as school leadership and modeling of the innovation influence whether a technology is adopted for use (Hall, 2010). To contribute to the complexity, the successful implementation of a technological innovation, such as VR in the classroom, also hinges on a teacher's ability to move through their stages of concern (SoC) related to the innovation (Loucks & Hall, 1979). An individual's initial concerns focus on personal issues, (stage 2 of 6 on a 7-point Likert scale), primarily related to "time, planning, and instructional practices" (Donovan et al., 2007, p. 274).

Preservice teachers (PSTs) require explicit strategies to integrate VR into primary and secondary classrooms (Ferdig et al., 2018). Students who produced a written summary of the lesson after each VR segment outperformed those who did not pause to document their learning (Parong & Mayer, 2018). Lee and Shea (2020) found that introducing PSTs to VR with the intent to develop curriculum increased PSTs self-efficacy, and PSTs' intention to use VR in their future classroom. However, not all PSTs are convinced they should integrate VR into their teaching. Unlike Donovan et al., (2007) measure of teachers' initial SoC, more research is needed to explore an individual's change in concern following the implementation of a new technology. To address the need for an in-depth understanding of "content and context design" (Lytras et al., 2016, p. 878) when using VR, the purpose of this study is to examine PSTs acceptance to use VR to plan science instruction. This research seeks to inform how teacher educators, school administrators, and digital designers can acknowledge and respond to individual teacher concerns to help schools more effectively integrate VR technology to support teaching and learning in the classroom.

1.1. VR in education

The term "virtual reality" has been used since the 1960's, but differences between immersive and non-immersive VR require clarification. Ferdig et al., (2018) define non-immersive VR as a computer-based environment or simulation (e.g., iPad, laptop, desktop computer), whereas immersive VR is when an individual perceives him or herself inside the proposed setting (e.g., head-mounted viewer). Our study aligns with the immersive VR definition, which we hereafter refer to as VR.

VR as an instructional tool is a relatively new concept that holds considerable promise for the educational technology community because it offers potential for growth in student learning (Hutchison, 2018) and can help students understand abstract models (Chen et al., 2020). Though there is a paucity of research on VR in teacher preparation, the study of VR is noted within some education settings. For example, middle-school social studies students in economically disadvantaged rural school districts were introduced to VR to measure differences in motivation and student achievement (Bowen, 2018). Similar, but using augmented reality, 51 elementary students used augmented reality to engage in hands-on science learning (Wang, 2020). Although Wang (2020) found no statistical significance between using augmented reality and e-books, qualitative data indicates the bright colors offered by multimedia was an affordance that increased student discussion.

In higher education, undergraduate medical students used VR to help learn about the structure of the human body, and results also found that immersion and imagination features of VR-mediated course content positively impacted perceived usefulness and perceived ease of use (Huang et al., 2016). Harron et al. (2017) also explored the design of a museum learning experience using virtual reality and mobile devices. To contribute to students' understanding of science concepts, a classroom of elementary learners used VR to access the View-Master National Geographic Wildlife app (Hutchison, 2018). Similarly, in another example, VR devices helped deliver instruction to students in remote locations to transform social interaction via behavior and context (Bailenson et al., 2008). Additionally, computer-based VR using a researcher-designed software known as Topo-Pano Explorer helped teach the concept of contour mapping by displaying immersive 360° images of real-world environments (Park et al., 2008). Though exploring VR implementation in the classroom is essential to determine strategies for using VR, aligning VR to standards-based instruction needs to be a curricular imperative.

VR is an effective mechanism to increase student engagement (Bowen, 2018; Ferdig et al., 2018; Merchant et al., 2014). Most recently, Bowen (2018) realized gains in student achievement and engagement when VR was implemented in a rural middle school to teach social studies. Merchant et al., (2014) examined 29 studies on virtual worlds in K-12 and higher education and concluded that, overall, virtual worlds positively affected student outcomes. VR-based instruction may also benefit English learners' language development (Craddock, 2018) and students with cognitive or psychological disorders (Horn, 2016). Despite some evidence to suggest that VR may benefit learning and increase student engagement, we still know little about how a teacher perceives and intends to use VR in the classroom.

1.1.1. Immersive virtual reality

Since 2016, low-cost VR devices and curriculum resources such as Google Cardboard head-mounted viewers and the Google Expeditions app ("Expeditions") have become widely available and accessible (Churchill, 2017). Google Cardboard is a VR head-mounted viewer intended to be used with a smartphone or small tablet (Google Cardboard, n.d.). Similar viewing devices are readily available from other manufacturers, but since Google introduced Google Cardboard for $15 USD, this viewer has garnered the attention of classroom teachers (Horn, 2016). Teachers can create their own immersive VR classroom kit for as little as $150 and download a variety of free apps (Long & Eutsler, 2020).

The Expeditions app is designed to support learning in K-12 classrooms (Dutton, 2016). Each expedition is a virtual tour of a location viewable by students with a VR head-mounted viewer. Locations include sites from outer space, the inside of an atom, and the caldera of a volcano. The teacher utilizes a tablet to direct students to specific points and can observe where students are focusing their attention within the tour (Google Expeditions, n.d.). Expeditions boasts over 900 virtual tours available for open access and use to allow students to see around the world from the safety of their desks (Ullman, 2016). Twenty-four elementary students attended a 2-week social studies camp, and results showed that after using the Expeditions app to engage in immersive VR learning, student motivation improved, coupled with a diminishment of test anxiety (Cheng & Tsai, 2019). Companies who sell VR, such as RobotLab, are developing VR Expeditions 2.0 in partner with Britannica, slated to be introduced in late 2021 (Galvis, 2020). As VR in education affordances continue to evolve, more research is needed to examine how to plan instructional experiences with VR.

2. Concerns-based adoption model

The concerns-based adoption model (CBAM) is a technology adoption model that identifies an individual's concerns about an innovation (Hall, 1976), such as new technology. This model offers an in-depth measure of an individual's concerns, as opposed to the diffusion of innovation theory (Rogers, 2003) which analyzes innovation acceptance over time. CBAM is comprised of three dimensions to understand technology acceptance: SoC, levels of use, and innovation configuration. When integrating a new technology, these concerns are influenced by the individual's background experiences, self-efficacy related to the technological innovation, and the intended purpose for introducing the target technology. When we implement the CBAM model and focus on the SoC dimension, it allows for the exploration of potential adopters perceived needs and perceptions relative to the innovation. Figure 1 displays the SoC about the innovation, measured on a seven-point scale that ranges from zero to six: awareness (0), informational (1), personal (2), management (3), consequence (4), collaboration (5), and refocusing (6). The CBAM acts as a guiding framework to help illuminate concerns that accompany PSTs perceptions and intentions to use VR. These illuminated concerns make it possible to tailor technology integration strategies and teacher professional development.

IMPACT	6	Refocusing	The individual focuses on exploring ways to reap more universal benefits from the innovation, including the possibility of making major changes to it or replacing it with a more powerful alternative.
	5	Collaboration	The individual focuses on coordinating and cooperating with others regarding use of the innovation.
	4	Consequence	The individual focuses on the innovation's impact on students in his or her immediate sphere of influence. Considerations include the relevance of the innovation for students; the evaluation of student outcomes, including performance and competencies; and the changes needed to improve student outcomes.
TASK	3	Management	The individual focuses on the processes and tasks of using the innovation and the best use of information and resources. Issues related to efficiency, organizing, managing, and scheduling dominate.
SELF	2	Personal	The individual is uncertain about the demands of the innovation, his or her adequacy to meet those demands, and/or his or her role with the innovation. The individual is analyzing his or her relationship to the reward structure of the organization, determining his or her part in decision making, and considering potential conflicts with existing structures or personal commitment. Concerns also might involve the financial or status implications of the program for the individual and his or her colleagues.
	1	Informational	The individual indicates a general awareness of the innovation and interest in learning more details about it. The individual does not seem to be worried about himself or herself in relation to the innovation. Any interest is in impersonal, substantive aspects of the innovation, such as its general characteristics, effects, and requirements for use.
	0	Unconcerned	The individual indicates little concern about or involvement with the innovation.

Figure 1. Stages of concern (SoC) about the innovation (George et al., 2006)

Adapted to the context of this study are examples of each SoC. An example of a concern at the awareness stage could be an individual admitting they have never used VR. Informational constitutes the information seeking stage where an individual may ponder what apps and VR tours are compatible with the VR hardware. In the next stage, personal, the concern might center on how VR influences an individual personally, which could include questioning if immersive VR might lead to dizziness or another unpleasant feeling. During the management level, the user might focus on integrating VR into lesson planning (e.g., selecting the appropriate app, VR tour, scaffolding with supportive learning activities) in addition to expressing concerns about how to maintain the VR viewers and smartphone devices (e.g., charging the devices, updating software, downloading VR tours). At the consequence stage, the teacher might inquire about how VR can influence student motivation, engagement, and achievement. Within the collaboration stage, concerns might focus on wondering how others are using VR compared to how the individual uses the innovation in their own classroom. The final stage, refocusing, occurs when the teacher seeks out new ways to improve the teaching and learning experience, which could entail searching for new and improved software (e.g., VR apps) and higher quality hardware.

Similar to other technology innovations, implementation barriers can stem from an individual's beliefs about technology, self-efficacy, reaction to change, availability of technology resources, lack of professional development, and inability to manage the tool (Hall, 1976). Despite the promise of VR's ability to improve student motivation and learning, few teachers integrate VR into their classrooms. With access to more affordable head-mounted viewers, it is important to investigate PSTs' attitudes, perceptions, and intentions to use VR in the classroom so teacher preparation programs can address and prepare for perceived barriers that might otherwise inhibit technology acceptance and use.

2.1. Research questions

- How do PSTs' SoC before the introduction of VR compare to their SoC following a hands-on intervention with the innovation?
- Relying on open-ended data sources (e.g., survey items, science journals, focus group), what appear to be the underlying reasons to explain each PSTs' concerns when using VR to plan science instruction?

3. Method

This mixed-method research study (Creswell & Creswell, 2018) provides empirical and explained perceptions to compare each participant profile with the innovation before and after exposure to VR. This method was most appropriate to help closely examine the sample size of five PSTs from the same undergraduate classroom. Multiple data sources provide triangulation to answer the study's research question. We implemented a pre and post design of the Stages of Concern Questionnaire (SoCQ), which contains 35 Likert-scale response items, to which we added one open-ended response item (George et al., 2006). Although the SoCQ reports quantitative data, the questionnaire produces qualitative descriptions based on the quantitative measures obtained. Other qualitative data sources include PSTs' individual science journals, individual surveys following each VR experience, and a focus group meeting to conclude the study.

3.1. Participants, context, and procedures

Data were collected in fall 2018 within an undergraduate science methods course at a large, public, Hispanic-serving and minority-serving institution, comprised of more than 38,000 students in the southwestern United States, with approximately half first-generation college students. PSTs enrolled in the course as part of their degree requirement toward becoming middle-school science teachers. Participants included Amber, Ava, Carla, Chuck, and Frieda (pseudonyms). Chuck and Frieda were student teaching full-time in grades five and seven, respectively; Amber, Ava, and Carla were observing in middle-grade classrooms twice a week as part of their teacher preparation. Three of the five participants came from diverse backgrounds; Table 1 displays the study participants reported gender, ethnicity, and age.

Table 1. Demographic makeup of study participants

Participant	Gender	Ethnicity	Age
Amber	Female	Caucasian	25
Ava	Female	Hispanic	24
Carla	Female	Hispanic	21
Chuck	Male	Caucasian	25
Frieda	Female	Asian	27

This study was introduced to PSTs as an opportunity to engage and explore with VR to innovate science learning. Participation had no impact on their course grade, and informed consent was obtained from each participant. The consent outlined the risks involved, including the potential for "loss of spatial awareness, dizziness and disorientation, seizures (e.g., anyone who is epileptic or is prone to seizures should not be considered as a candidate for VR), nausea, and eye soreness."

Four of the five VR sessions were conducted in a science education laboratory classroom. The classroom, equipped with high top laboratory tables with four swivel laboratory chairs at each table also contained sinks, and most equipment typically found within a middle-school science lab. One intervention session was held at the university's environmental education center. The environmental education center contains indoor and outdoor areas designed for informal learning. The outdoor learning area includes two paleontology/archaeology dig boxes, a Texas geologic history walk, and a pond suitable for practicing environmental monitoring and collection of macroinvertebrates.

The VR equipment used in this study was a commercially available classroom set containing 20 headset viewers, a teacher tablet, 360-degree camera, and LAN router. The viewers consisted of a hard-plastic shell with magnifying lenses that allowed for viewing VR content displayed on the included smartphone. Unlike the more affordable cardboard viewers, these viewers are sturdier and include a focus adjustment. The viewers did not come with a head strap, which required students to hold the viewer when experiencing the VR lesson.

3.2. Data collection

This study was conducted over five consecutive weeks. Before PSTs were introduced to VR, they completed the SoCQ, which contained the standard 35 items and demographic items. One additional open-ended response item concluded the survey, "Please share anything else you would like to tell us about your experience/opinions regarding technology." Other qualitative data sources included individual science journals, individual surveys following each VR experience (n = 5), and a focus group meeting at the end of the study.

The SoCQ questionnaire "was designed for and is intended to be used strictly for diagnostic purposes for personnel involved in the 'adoption' of a process or product innovation" (Hall et al., 1977, p. 57). The questionnaire contains 35 statements, with each focused on a specific concern in connection to the innovation. Respondents indicate the degree to which each concern is true for them by selecting from a 0–6 Likert-scale. High numbers indicate high concern, low numbers are associated with low concern, and 0 indicates very low concern or irrelevant items (George et al., 2006, p. 26). Sample items include: "I am concerned about students' attitudes toward the innovation," "I have a very limited knowledge of the innovation," and "I would like to know how the innovation is better than what we have now."

Details of the intervention illuminate PSTs depth of involvement with the VR. Each 90-minute session built on the previous experience to move at a pace dictated by the participants (Pearson et al., 2019), with lessons intended to help PSTs' envision VR in their science classrooms.

We introduced the study and PSTs completed the SoCQ (Hall et al., 1977). To administer the survey, we loaded the original questionnaire items into Qualtrics and PSTs completed the electronic survey on their personal laptops. Following the survey, we unboxed the VR and provided basic instructions on how to power on and handle the equipment. The learning objective during session one was for PSTs to become familiar with the VR device. To accomplish this goal, we directed PSTs to the functions of the hardware and took them on a VR tour, using the Expeditions app. For the second VR experience, PSTs compared Expeditions tours designed by different developers intended to teach life science. Though most VR tours are silent and only include on-screen text, PSTs went on an Expedition tour that was equipped with audio features, audible through the teacher tablet. We integrated individual science journals that contained guiding instructional questions to support and encourage responses and reflective thinking, which were adapted from the Expeditions curriculum.

Within the third VR experience, PSTs' aligned science-learning standards with VR videos, while continuing to integrate science-learning journals. To accomplish this, PSTs explored the Discovery VR and roller coaster apps and brainstormed how these immersive videos might influence students' personalized learning experience and understanding of the standard. Additional apps explored were Titans of Space Cardboard VR, SciVR, Share the Science: STEM, and Water Cycle VR. By the fourth VR session, we perceived PSTs were ready to create their own VR tour. Using a Ricoh Theta SC 360° camera, they took turns taking photos of an outdoor science center located on the university campus. Some of the photographed settings included a fossil dig, pond, nature trail, waterfall, and life-size reconstructions of extinct animals. Photos from the 360° camera were transferred to a desktop computer using Google Poly, a free VR tour creator software. The self-made VR tour was made publicly available online at [link available upon acceptance]. Each PST typed the web address into the browser on their VR smartphone device, placed the phone into their head-mounted viewer, and immersed themselves into the VR tour they had created. Again, PSTs responded to question prompts and documented their thinking in their science journals. At the completion of the study, PSTs completed the SoCQ for the second time.

3.3. Data analysis

The SoCQ was analyzed in accordance with the guidelines outlined in the questionnaire manual (George et al., 2006). The scores for each of the SoCQ stages were averaged and the resulting means were multiplied by five to produce a raw score. The raw score for each stage was converted into a percentile score utilizing the tables derived by George et al., (2006). As illustrated in the graphs, the resulting percentile scores were graphed with a description to produce an individual profile for each participant. Each PSTs result was matched to one of these profiles to form the basis for the descriptions of their SoCs.

Guided by Braun and Clarke (2006), we engaged in a seven-phase deductive thematic analysis of open-ended data (science journals, pre and post SoCQ open-ended item, surveys, focus group). During phase one, we individually read the complete qualitative dataset, documenting impressions, thoughts, and preliminary interpretations. Following our initial readings, phase two consisted of rereading all open-ended responses to deductively align data in accordance with each CBAM SoC level (Hall, 1976). This deductive process allowed for a deeper understanding to compare each participant's SoC at the beginning and end of the study. In phase three, we collaborated, compared, and contrasted our sorting of open-ended data to align with each SoC, to deepen our understanding of each participant's experience. Phase four involved individually delving back into the dataset to determine whether all data fit within the deductive SoC domains, ascertain whether data were sorted accurately, and distinguish if new domains were needed. After sorting the data, our fifth and final phase served to ensure a shared understanding of each theme.

To ensure inter-rater reliability, the data was independently analyzed by both authors and discussions were held until consensus was reached. To address issues of validity, we employed member checking by returning the data and our interpretations to the participants to allow PSTs to support, explain, corroborate, or clarify their responses. To further address validity, we divulge the role of the researchers. Both researchers were involved in the design, implementation, and analysis of the present study. The primary researcher was an experienced educational technology user, but had never used VR head-mounted viewers prior to this study. The co-researcher has a background in computer science, was the science methods professor, and had previously used VR to access Google Expeditions in his 8th-grade classroom.

4. Results and discussion

Given the potential benefits of using immersive VR to support learning, such as increased student engagement (Bowen, 2018; Ferdig et al., 2018; Merchant et al., 2014) and opportunities for inclusion (Craddock, 2018; Horn, 2016), it is important to understand what barriers to implementation are perceived by teachers who consider incorporating VR into their instruction. Since implementation of VR in the classroom may represent a large investment of teachers' and school districts' budget allocation, it is important for schools and districts to understand barriers that could otherwise hinder a valuable return on their investment. As outlined in George et al., (2006) SoCQ manual, findings are organized and discussed by each participant's individual peak stage score interpretation before and after the study, and relative to their individual profile interpretation description.

4.1. Individual peak stage score interpretation

Peak stage score interpretation is the simplest form of SoCQ data analysis. To perform this analysis, we charted the percentile scores and identified the highest stage of concern. In cases where another stage was within one or two points, both stages were identified as the highest stage of concern. It is important to note that percentiles are not interpreted as absolute scores; rather, they are viewed as relative to the other stages. Individual second highest scores were also examined to identify a secondary SoC.

4.1.1. Before the VR experience

Table 2 shows the individual peak scores for each participant prior to the study. Stage 0 (unconcerned) was the peak score for both Amber and Carla, indicating that teaching and learning with VR was not necessarily a high priority. Interestingly, unique to both Amber and Carla, their second highest SoC was stage 1 (informational) suggesting that while not a high priority for them, they were interested in learning about VR.

For Ava and Frieda, their highest SoC was stage 2 (personal), which suggests they were unsure about the demands of implementing VR in science. Akin to Amber and Carla, both Ava and Frieda shared Stage 1 as their second highest score, which indicates that although they had concerns about using VR to support instruction, they were interested in learning more about the innovation.

Chuck's individual peak scores in stages 1 and 2 are nearly identical, indicating a profile similar to Ava and Frieda. However, he also has a relatively high score in Stage 0, which invites a comparison to Amber and Carla. It is interesting to note that there appear to be two subgroups, Amber–Carla and Ava–Frieda, with Chuck being a hybrid of both groups.

Table 2. Individual peak stage score interpretation pre

Participant	S0	S1	S2	S3	S4	S5	S6
Amber	99	*80*	48	39	69	16	5
Ava	31	*80*	83	56	24	28	17
Carla	99	*95*	89	60	43	76	34
Chuck	81	*96*	97	34	82	93	42
Frieda	31	*54*	59	30	11	10	17

Note. Peak scores are underlined.

4.1.2. After the VR experience

Table 3 displays the individual peak scores after the study ended. Several interesting observations appeared when analyzing the post-study scores. For one, the mini-cohort of Amber and Carla showed progress. While they both maintained Stage 0 as their primary SoC, their second highest SoC shifted upward on the model's trajectory. Amber moved from the information stage into Stage 3 (management), which indicates she is contemplating how to implement VR in her classroom. Carla shifted from Stage 1 to Stage 2, suggesting that she learned about the innovation but is now concerned with her ability to implement VR in the classroom.

The mini-cohort of Ava and Frieda indicated stagnant change. Both showed Stage 0 among their primary SoC with stages 1 and 2 nearly as prominent. This profile shows that they are still struggling with their self-efficacy, and the idea of implementing technology is no longer a high priority.

Chuck, who shared similarities with both mini-cohorts, had the most dramatic change in peak scores. Chuck's post-study peak score is at Stage 5 (collaboration), which indicates that he is focused on connecting with others to use the technology in more innovative ways. Chuck's second highest peak scores were noted in stages 1 and 2, indicating his concern with personal implementation and a desire to learn more.

Table 3. Individual peak stage score interpretation post

Participant	S0	S1	S2	S3	S4	S5	S6
Amber	99	45	7	*80*	5	5	11
Ava	*81*	80	83	60	16	40	38
Carla	98	*84*	92	73	43	59	57
Chuck	31	91	*92*	43	22	97	26
Frieda	81	*60*	57	52	19	22	30

Note. Peak scores are underlined.

4.2. Individual profile interpretation

The richest method for interpretation of the SoCQ occurs through individual profile analysis (George et al., 2006). Profile analysis is accomplished by graphing the SoC percentile scores for each participant. Ideally, peaks in the profile should move from left to right as the participant progresses into higher stages of acceptance. This movement was only noted among some of the study participants. To facilitate discussion of the individual profile interpretations, each student profile is presented as a separate figure. A frequency of the highest stage of concern is displayed to show collective concerns before (Table 4) and after the study (Table 5).

Table 4. Frequency of highest concerns stage for the individuals pre

Participant	S0	S1	S2	S3	S4	S5	S6
Number of participants	2	1	3	0	0	0	0
Percent of participants	33	17	50	0	0	0	0

Table 5. Frequency of highest concerns stage for the individuals post

Participant	S0	S1	S2	S3	S4	S5	S6
Number of participants	4	0	1	0	0	1	0
Percent of participants	67	0	17	0	0	17	0

4.2.1. Ava

Figure 2 displays the profile for Ava before and after the VR experience. Both pre and post lines of Ava's profile closely resembles what George et al., (2006) describe as a typical non-user profile. This profile suggests that while Ava is interested in learning more about the innovation, she is not overly concerned with other stages, such as management and collaboration. Although both lines adhere to the non-user profile, the increase in stage 5 and 6 (refocusing) scores indicate that Ava has become more involved in thinking about collaborating and/or is refocusing with competing ideas. The difference in the Stage 0 score is not as important in this profile configuration as are the stage 1 and 2 scores, because earlier conclusions have shown that "variations in stage 0 do not seem to be as important as variations in stage 1 and 2" (George et al., 2006, p. 38).

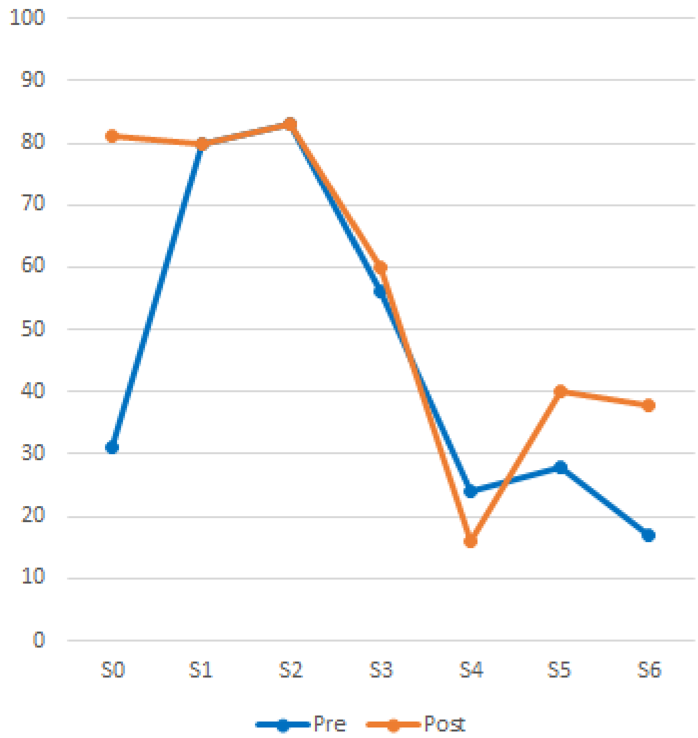


Figure 2. Ava's pre and post experience profile

4.2.2 Frieda

Figure 3 shows Frieda's pre and post experience profiles. Frieda exhibited a variation on the typical non-user profile, described by George et al., (2006) as a negative one-two split with tailing up at stage 6. A negative one-two split occurs when stage 2 is markedly higher than stage 1. In Ava's case, stage 2 was higher but only by a couple of points so it would not classify as a negative split. A negative one-two split indicates potential resistance to the use of VR in the classroom due to elevated personal concerns to use the technology. Users with this profile will not typically move toward adoption until stage 2 concerns are addressed (George et al., 2006). Frieda's post profile line removes the negative one-two split, and although the later stage of concerns are elevated, this implies that Frieda has started to move toward the adoption of VR. The increase in stage 1 indicates that she likely wants to learn more before exploring the innovation further.

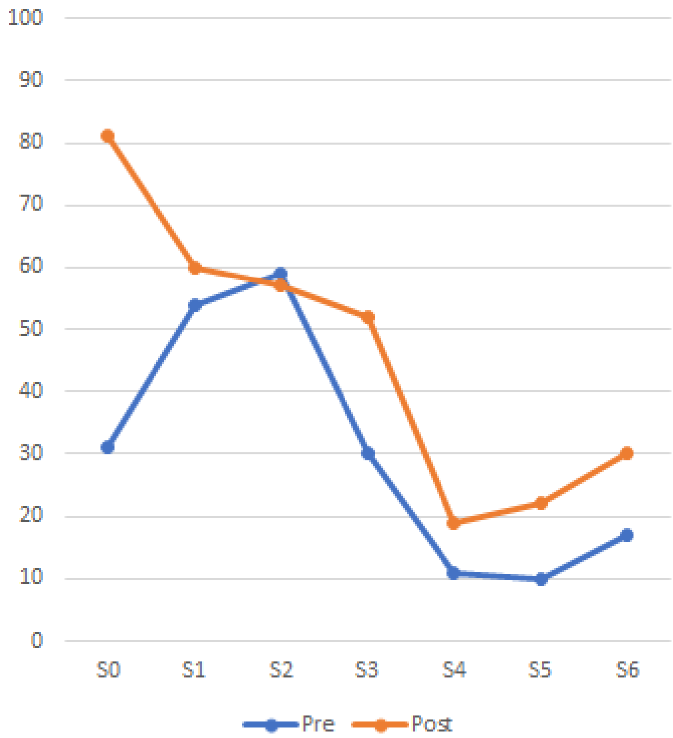


Figure 3. Frieda's pre and post experience profile

4.2.3. Chuck

Chuck's profile (Figure 4) also shows signs of a typical non-user but with the added characteristics of a high collaboration and consequence concerns profile. The high score in both stage 4 and 5 show that Chuck was initially concerned with how implementing VR would impact his students, but he was also interested in collaborating with others to use the technology. The drastic drop in stage 4 after the VR experiences indicates that Chuck is confident that his students would benefit from VR. However, Chuck still has concerns associated with the non-user profile, which indicates he would like to seek additional information.

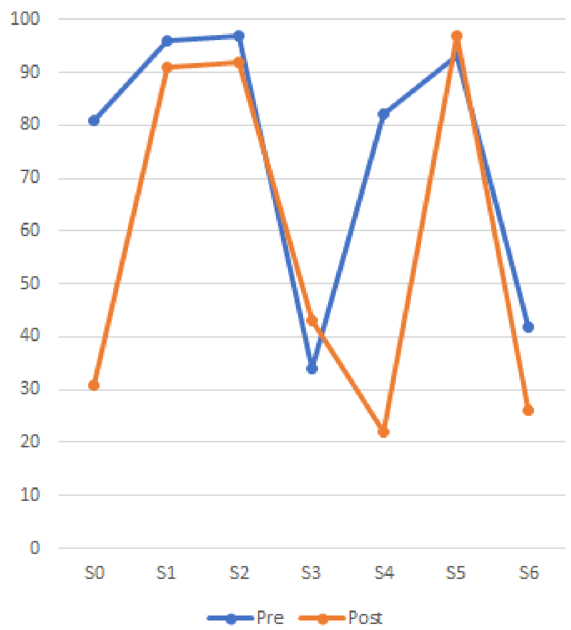


Figure 4. Chuck's pre and post experience profile

4.2.4. Carla

Figure 5 exhibits Carla's pre and post profiles. Carla began the VR experience as a typical non-user profile as described by George et al., (2006). At first glance, it may not seem like much change from pre to post; however, a closer examination reveals a negative one-two split similar to that portrayed in Frieda's pre-experience. This indicates that Carla's concerns about how to implement VR affects her personally, which now overrides her desire to learn more about the topic. It is interesting to note the reduction in Carla's stage 5 concern. This reduction suggests that she is more confident in collaborating with colleagues to use VR in the classroom. A stage 6 increase suggests she may be exploring a more powerful alternative and/or different ways to use the tool.

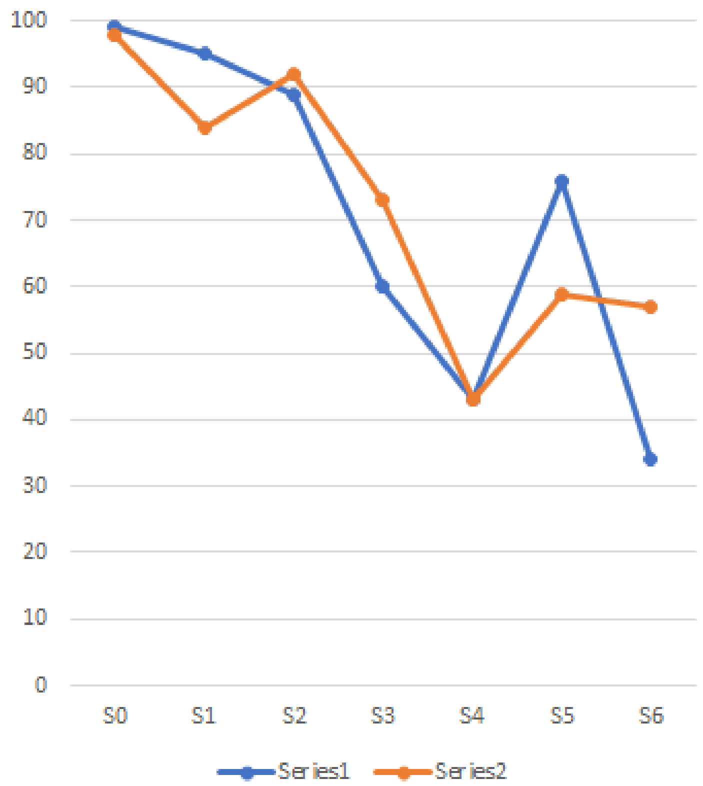


Figure 5. Carla's pre and post experience profile

4.2.5. Amber

Amber's pre and post profiles are depicted in Figure 6. Amber's initial profile indicates a high interest in stage 1 and a secondary concern in stage 4, suggesting that she is curious about VR but has reservations about what it means for her students. Amber's post experience profile shows that she has possibly lost enthusiasm for the innovation. The high stage 3 score could imply that she thinks the implementation of VR would require too much commitment in terms of time, logistics, and/or management.

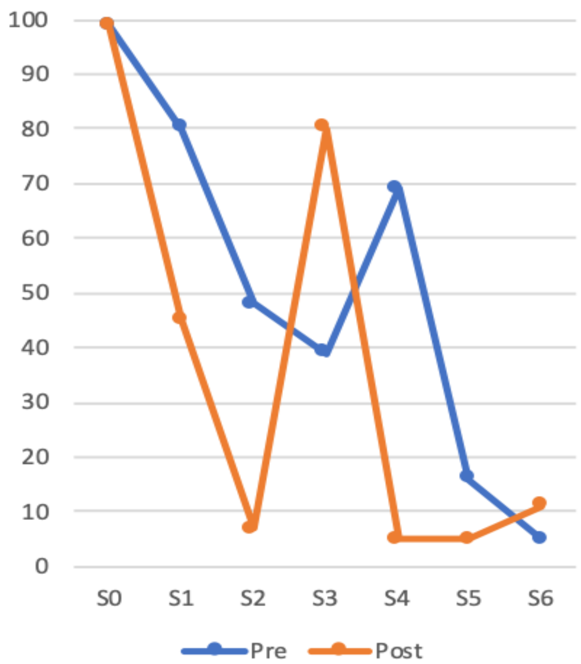


Figure 6. Amber's pre and post experience profile

4.3. Collective reflection of PSTs VR concerns

After unpacking PSTs concerns, open-ended response data revealed four additional themes: emphasis on the technical aspects of VR, learning engagement/satisfaction, generation of lesson plan ideas, and PSTs intention to use VR.

4.3.1. Emphasis on the technical aspects

Considering the technical properties of the VR headset, PSTs judged device quality and proposed suggestions. Four explained the clarity of the VR headsets was a problem. Carla said her main concern was quality and that it "would be great if the quality was better...that is my main concern." Chuck wished "there was some more clarity on the images." Frieda offered a suggestion to "try looking into if the blurriness came from goggles or the device," and explained that "you would have to have a better-performing device/cellphone than the provided Nokia," because "the Nokia was blurry and without finding a hot-spot, it would not be seen in good quality." Amber elaborated on the eyepiece and explained a possible solution to reduce motion sickness while wearing the VR headset, because "resolution is blurry possibly because of the eyepiece. This may help with reduction of motion sickness."
Ava noticed a difference among the quality of images and how they vary from one VR tour developer to another. For example,

I felt like the Life in the Deep Ocean expedition lost my attention because the images were fake and I could tell and I feel as though that would be the time to just pull images up on the board and have the students look versus using VR because it's not 'real' like the other images were.

Chuck offered a possible explanation for the differences among the image quality, saying that "the illustrations were okay- I understand that its purpose being that those images are harder to capture, and that some things must be illustrated for VR." Chuck added the importance of using a high-quality smartphone, "once the quality of the phone was changed (to Samsung™ Galaxy S7), I was really amazed at how great the experience became!" explaining, "I know I was being tough on it, but I'm just putting myself in the minds of my students and thinking what they would think."

Two students suggested head straps. Carla said that she felt she would "need straps to wear around my head" and Chuck said, "I wish there were straps on the headsets so I wouldn't have to hold up the goggles the whole time." Those who wore glasses faced another challenge. Carla noted "it's a little hard to use the equipment with glasses but because I can still see without my glasses, I'm fine. I would be concerned about students who cannot see too well without their glasses though."

There was mention of the time it took to prepare the VR headsets for use. Frieda said that "creating our own VR tour helped us to understand how time consuming it is to set up everything." Frieda went on to speak about connectivity issues, saying that it was "easier to use and follow along as long as the Wi-Fi is properly working. It would be easier without technological conflicts." Another shared an idea that could make the experience more realistic. Ava said, "I wish we could record videos or switch the view to 1st person like in a go pro that way it really feels like you're there."

4.3.2. Learning engagement/satisfaction

Four PSTs said they enjoyed the experience of using VR. Chuck admitted that he "really enjoyed this experience. I think it is awesome that we got to go to different continents to see the different volcanoes." Frieda reminisced about being able to ride on a roller coaster, an otherwise unimaginable experience. She explained, "I also tried the Discovery app, the roller-coaster video in thrills and adventure. That was a good experience: it actually felt like being in a roller-coaster." Carla relayed that VR can give students an opportunity to visit places they would never get to see otherwise. "I think it was great, especially because the deep ocean is an environment that students would probably never see in real life." Carla also found it "impressive that we can make our own [VR tours]." Ava, though she liked the VR, felt that the VR research detracted from her teacher preparation experience.

I enjoyed this experience overall. Although I do feel a little like it took away from the methods class and us actually learning things that are pertainable [sic] to learning methods content I enjoyed it. It made class a little more fun in being able to mess around with the VR technology.

PSTs also liked the variety among the VR tours available. Frieda said, "Using the different expedition types gives it variety and different perspectives." Chuck enjoyed going to "different continents to see the different volcanoes," while Ava "loved the persevered oceans." Two other PSTs emphasized how realistic the experience felt for them. Ava admitted she "really enjoyed having the sound even if it was only from the teacher's device. I feel like that does add to the expedition and makes it more realistic." Frieda provided an example of realism, "between the Discovery VR app and the YouTube® video, the YouTube video made a difference visually and in audio." Frieda also noticed "having realistic scenes will allow students to experience an outside classroom experience."

Issues of equity and access arose in regard to the importance of taking future students to places they may not be able to afford to travel. Ava admitted for herself that money was an impeding factor, because "both places we visited are places I have wanted to go in real life but just don't have the money for yet." Carla seemed relieved that VR allowed for some travel adventure, "I think it was great, especially because the deep ocean is an environment that students would probably never see in real life." Frieda echoed, "The illustration allows students to see what they cannot experience as there are places where people can't visit."

4.3.3. Generation of lesson plan ideas

Regarding science lesson plan ideas, PSTs had multiple recommendations for how to incorporate VR into learning. Frieda proffered, a

field experience involving different terrains and areas dealing with succession. Life Science involves the outside world so an explanation of microhabitats and biodiversity could make a great experience. Areas such as creeks and forests could make a good lesson with the VR tour.

Some PSTs were inspired by the Elm Fork VR tour they created. Amber envisioned how "you could take a 360 photo of a specific biome and have the students list all the biotic and abiotic factors that they see." Carla had an idea to take "a picture of an area where we have all these factors and then make students list out the different factors" and then ask students to consider how those factors could "affect the relationships that organisms have with their environment." Ava adapted the idea to make it here own,
one VR tour I could do would be similar to the elm fork where I could go to a nature reserve or park and use that over a lesson about organisms and the environment. Another VR tour I could do would be to make a trip out to

the different ecoregions of Texas and take pictures so students can have an experience learning about the effects of weathering, erosion and deposition of the earth.

Chuck shared many different science lesson plan ideas, one involving students who might otherwise be unable to attend a field trip. He reflected, “I could use VR to simulate my trip for the students who didn’t go so they don’t totally miss out on the experience.” Additionally, Chuck came up with lesson plans to explore the solar system, different volcanoes, as well as parts of a cell for biology. Carla also had an idea to create a VR lesson plan for her students following a storm or natural disaster. She explained how this would be helpful, “so that we can learn about natural disasters or certain events. I liked creating our own experiences so I think students would definitely enjoy it too.”

4.3.4. Intention to use VR

In line with PSTs who were satisfied with the VR experience, four PSTs intend to use VR in their classroom because they were positively impacted by the experience. Frieda reflected, “it’s a great way to have students experience their surroundings especially if there are students that have not experienced a 360 POV. It’s really a good idea to use VR as a classroom experience.” Mindful of the limitations of available technology in schools, Chuck said, “I’d use VR in my classroom if I had it! I think VR experiences can be a powerful thing for teachers and students.” Chuck also admitted the need for additional professional development, “I would like to do VR experiences with my students one day, given a little more training to master the technology.” Carla added that, “it’s also cool to think that teachers can create their own experiences for their classes.” Ava appeared impressed with the ability to make her own VR tour, “it’s cool to see the finished piece of what we did last week.” Carla believed that by using VR, “it would engage students to be able to create their own VR experiences.”

Both Frieda and Carla explained how the experience could help engage their students and improve their lesson plans. Frieda said, “I think it would be great to use in the classroom and making our own with experience we want to share with our students could further better our lessons.” Ava believed VR would be helpful because “students would really enjoy going out and taking the pictures and then being able to see it through a different lens.”

4.4. Reflection of the intervention experience

To inform future research, we share lessons learned from this intervention experience. Logistics pertain to selection, planning, and implementation. First, it is essential to identify the usability experience. Decide on the desired VR interface, whether students will use a desktop computer, laptop, tablet, or head-mounted viewer. Next, consider when and how the devices will be utilized during instruction because device options vary in quality and movability. For example, the Oculus Rift® is high quality but requires a wired connection, whereas the smartphone and head mounted viewer setup is wireless. A wireless education kit typically comes equipped with a secure and padded carrying case, 10 headset viewers, 10 smartphones, charging ports, and a router, but it is costlier. A third option is to create a set of five for a classroom for about $150, which might contain the head-mounted viewers with a strap, gently used but higher quality smartphones, with an option for teacher-led instruction using a tablet and router. Add a 360° camera to help students create their own VR experiences. Device selection is important because it involves careful consideration of intended use, budget, and subsequently, quality.

With devices in hand, device exploration is foundational to using apps and aligning with instruction. Just as teachers need to preview the curriculum before designing a lesson, it is essential to practice device setup. Spend time locating or designing VR tours to support instruction. Prepare to introduce the innovation to students. Thus, adapt classroom management techniques, emphasize safety procedures, and provide students with a laminated printout with access, use, charging, and cleaning instructions. Finally, obtain student feedback to adapt instruction, brainstorm new ways to engage with VR, and challenge students to design their own VR tours.

4.5. Limitations

We acknowledge that this study contains limitations. First, the sample size of five participants is small, and therefore cannot be generalized to other contexts. Nonetheless, the rich dataset is triangulated by open-ended response items, and the SoCQ provides descriptive behaviors to explain acceptance to a technological innovation. Another limitation is that data from the SoCQ and open-ended responses are self-reported, which may include social desirability bias (Edwards, 1957), where participants' could portray themselves as a more positive or confident user than is true.

5. Implications and future research

Implications of this research first indicate the need to extract an individual's concerns before implementing a technology, because an awareness of an individual's SoC could reduce barriers toward the acceptance of an innovation. Our findings also demonstrate the importance of providing PSTs with hands-on technology experiences, which is essential for an individual to conceptualize and actualize themselves using a new technology and is a necessary process that supports PSTs becoming classroom teachers. In our study, we noticed that all PSTs perceived and responded differently to using VR to support learning, in addition to how they envisioned using technology in their future classrooms. Although each PST had a unique profile, three subgroups emerged in the data, which suggests that professional development should align with a user's individual SoC profile. Therefore, we recommend additional research examines how larger subgroups of PSTs demand unique supports as they move through their SoCs.

A longitudinal and more expansive approach similar to this research design could help distinguish the amount of time required by the individual to reach the two highest stages of concern, collaboration and refocusing. With regard to in-service teachers, researchers might consider investigating how these practicing teachers perceive the usefulness of using immersive VR in the science classroom over the long-term. This research could also focus on improving our understanding of K-12 teachers' willingness to integrate VR and how their students perceive and perform when learning with VR. To expand our understanding of how to help teachers use VR in the classroom, case studies might investigate teachers who are unfamiliar with VR and exhibit resistance toward adopting the technology. With comprehensive data collection via a large-scale research design, this could lead to the inception of a technology integration preparation manual for administrators, teachers, and teacher educators implementing VR.

To contribute to research on K-12 student achievement when learning with VR, we recommend researchers conduct a quasi-experimental study to measure the use of VR in the classroom on students' learning outcomes. We also encourage a study on the use of VR to measure students' self-efficacy in science before and after integrating VR as a tool to support learning. Finally, we urge researchers to explore the possibilities of VR in special education, to identify ways VR could support students with disabilities, such as taking them to locations that would otherwise be inaccessible.

6. Conclusion

The purpose of this study was to further examine the intersection of technology, design and pedagogy (Ioannou & Ioannou, 2020) when using VR. Therefore, by exposing PSTs to VR (Long & Eutsler, 2020) with the intention of expanding on the science curriculum and content (Lytras et al., 2018), we measured PSTs concerns to integrate VR when planning science instruction. Framed by the CBAM (Hall, 1976), we measured the SoC using the SoCQ before and after a 5-session hands-on exposure to the innovation. During these sessions, PSTs became familiar with the hardware, available VR apps, and brainstormed ways to connect science content to the experience of using and designing VR tours. An analysis of PSTs SoC shows that while four out of five participants intend to use VR within their future science classrooms, they desire more professional development and an opportunity to use VR with middle-grades students before they would implement VR with their own students. Open-ended data also found that PSTs focused on improving the technical aspects of VR, most were satisfied and enjoyed designing instruction and learning with VR, and these hands-on experiences of creating their own tour and engaging with VR led to the generation of multiple lesson plan ideas. This study provides a beginning model to introduce VR within teacher preparation programs and with middle-grades science teachers. Future research might replicate and expand this research design to include the investigation of in-service teachers and middle-grades students using VR to support science instruction.

Acknowledgements

We thank the preservice teachers who participated in this study. We also thank Carol Wickstrom, Ph.D. and Janelle Mathis, Ph.D., for their careful critique to improve this manuscript.

References

Bailenson, J. N., Yee, N., Blascovich, J., Beall, A. C., Lundblad, N., & Jin, M. (2008). The Use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *The Journal of the Learning Sciences*, *17*(1), 102-141. doi:10.1080/10508400701793141

Bowen, M. M. (2018). *Effect of virtual reality on motivation and achievement of middle-school students* (Unpublished doctoral dissertation). The University of Memphis, Memphis, TN.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101. doi:10.1191/1478088706qp063oa

Chen, J. C., Huang, Y., Lin, K. Y., Chang, Y. S., Lin, H. C., Lin, C. Y., & Hsiao, H. S. (2020). Developing a hands-on activity using virtual reality to help students learn by doing. *Journal of Computer Assisted Learning*, *36*(1), 46-60. doi:10.1111/jcal.12389

Cheng, K. H., & Tsai, C. C. (2019). A Case study of immersive virtual field trips in an elementary classroom: Students' learning experience and teacher-student interaction behaviors. *Computers & Education, 140,* 103600. doi:10.1016/j.compedu.2019.103600

Churchill, D. (2017). Emerging possibilities for design of digital resources for learning. In *Digital Resources for Learning* (pp. 227-246). doi:10.1007/978-981-10-3776-4_10

Craddock, I. M. (2018). Immersive virtual reality, google expeditions, and English language learning. *Library Technology Reports, 54*(4), 7-9.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Thousand Oak, CA: Sage.

Donovan, L., Hartley, K., & Strudler, N. (2007). Teacher concerns during initial implementation of a one-to-one laptop initiative at the middle school level. *Journal of Research on Technology in Education, 39*(3), 263-286. doi:10.1080/15391523.2007.10782483

Dutton, L. (2016). Google expeditions. *The School Librarian, 64*(3), 147-147.

Echenique, E. G., Molías, L. M., & Bullen, M. (2015). Students in higher education: Social and academic uses of digital technology. *International Journal of Educational Technology in Higher Education, 12*(1), 25-37. doi:10.7238/rusc.v12i1.2078

Edwards, A. (1957). *The Social desirability variable in personality assessment and research*. The Dryden Press.

Ferdig, R. E., Gandolfi, E., & Immel, Z. (2018). Educational Opportunities for Immersive Virtual Reality. *Second Handbook of Information Technology in Primary and Secondary Education*, 1-12.

Galvis, N. (2020). *Introducing RobotLAB VR expeditions V2.0!* Retrieved from https://www.robotlab.com/blog/introducing-robotlab-vr-expeditions-v2.0

Gasaymeh, A. (2018). A Study of undergraduate students' use of Information and Communication Technology (ICT) and the factors affecting their use: A Developing country perspective. *EURASIA Journal of Mathematics, Science and Technology Education*, *14*(5), 1731-1746. doi:10.29333/ejmste/85118

George, A. A, Hall, G. E., & Stiegelbauer, S. M. (2006). *Measuring implementation in schools: The Stages of concern questionnaire* (Appendices A-C, pp. 77-91). Retrieved from http://sedl.org/cbam/socq_manual_201410.pdf

Google Expeditions. (n.d.). Retrieved from https://edu.google.com/products/vr-ar/expeditions/?modal_active=none

Google Cardboard. (n.d.). Retrieved from https://vr.google.com/cardboard/

Hall, G. E. (2010). Technology's achilles heel: Achieving high-quality implementation. *Journal of Research on Technology in Education*, *42*(3), 231-253. doi:10.1080/15391523.2010.10782550

Hall, G. E., George, A. A., & Rutherford, W. L. (1977). *Measuring stages of concern about the innovation: A Manual for the use of the SoC questionnaire.* Austin, TX: Research and Development Center for Teacher Education, The University of Texas.

Hall, G. E. (1976). The Study of individual teacher and professor concerns about innovations. *Journal of Teacher Education, 27*(1), 22-23. doi:10.1177/002248717602700106

Harron, J., Petrosino, A. & Jenevein, S. (2017). Development of museum learning experience using virtual reality and mobile devices. In P. Resta & S. Smith (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference* (pp. 1592-1597). Austin, TX, United States: Association for the Advancement of Computing in Education (AACE). Retrieved from https://www.learntechlib.org/primary/p/177441/

Horn, M. B. (2016, August 17). *Virtual reality disruption* [Audio podcast]. Retrieved from https://www.educationnext.org/virtual-reality-disruption-3d-technology-education/

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. doi:10.1080/10494820.2013.817436

Hutchison, A. (2018). Using virtual reality to explore science and literacy concepts. *The Reading Teacher*, *72*(3), 343-353. doi:10.1002/trtr.1720

Ioannou, M., & Ioannou, A. (2020). Technology-enhanced embodied learning: Designing and evaluating a new classroom experience. *Educational Technology & Society, 23*(3), 81–94. doi:10.2307/26926428

Lee, C. K., & Shea, M. (2020). Exploring the use of virtual reality by pre-service elementary teachers for teaching science in the elementary classroom. *Journal of Research on Technology in Education*, *52*(2), 163-177. doi:10.1080/15391523.2020.1726234

Long, C. S., & Eutsler, L. (2020). Engaging with VR: Where will you take your students? *Science Scope, 43*(9). Retrieved from http://digital.ipcprintservices.com/publication/?i=663544&article_id=3698051&view=articleBrowser&ver=html5

Loucks, S. F., & Hall, G. E. (1979, April). *Implementing innovations in schools: A Concerns-based approach*. Paper presented at the American Educational Research Association Annual Meeting, San Francisco. (ERIC Document Reproduction Service No. ED 206 109)

Lytras, M. D., Damiani, E., & Mathkour, H. (2016). Virtual reality in learning, collaboration and behaviour: content, systems, strategies, context designs. *Behaviour & Information Technology*, *35*(11), 877-878. doi:10.1080/0144929X.2016.1235815

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A Meta-analysis. *Computers & Education*, *70*, 29-40. doi:10.1016/j.compedu.2013.07.033

Park, J., Carter, G., Butler, S., Slykhuis, D., & Reid-Griffin, A. (2008). Re-dimensional thinking in earth science: From 3-D virtual reality panoramas to 2-D contour maps. *Journal of Interactive Learning Research, 19*(1), 75-90. Retrieved from https://www.learntechlib.org/p/21831/

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology*, *110*(6), 785–797. doi:10.1037/edu0000241

Pearson, P. D., McVee, M. B., Shanahan, L. E. (2019). In the beginning: The historical and conceptual genesis of the gradual release of responsibility. In M. B. McVee, E. Ortlieb, J. Reichenberg, J., & P. D. Pearson (Eds), *The gradual release of responsibility in literacy research and practice* (pp. 1-21). doi:10.1108/S2048-045820190000010001

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed). New York, NY: Free Press.

Ullman, E. (2016, April 28). *The Next dimension: VR and AR in schools.* Retrieved from https://www.techlearning.com/resources/the-next-dimension

Wang, Y. H. (2020). Integrating games, e-books and AR techniques to support project-based science learning. *Educational Technology & Society, 23*(3), 53–67.

Understanding Pre-Service Teachers' Mobile Learning Readiness Using Theory of Planned Behavior

Kibar Sungur-Gül[1] and Hüseyin Ateş[2*]
[1]Department of Science Education, Nevşehir Hacı Bektaş Veli University, Nevşehir, Turkey // [2]Department of Science Education, Kırşehir Ahi Evran University, Kırşehir, Turkey // k.sngr89@gmail.com // huseyinates_38@hotmail.com
[*]Corresponding author

(Submitted September 26, 2020; Revised October 21, 2020; Accepted January 22, 2021)

ABSTRACT: The study aimed to understand pre-service teachers' mobile learning readiness with Theory of Planned Behavior using external salient beliefs. There were nine hypotheses tested with a total of 533 pre-service teachers in two cities in Turkey. Several scales adapted from Cheon et al. (2012) included 10 psychological variables. The results indicated that the TPB model explained 58% of the variance in intention to adopt mobile learning. The results of structural equation modeling (SEM) showed that the proposed model of the current study has acceptable fit data. The results of the SEM revealed that attitude, subjective norm and perceived behavioral control have significant impact on intention to adopt mobile learning. In addition, salient beliefs had an influence on the constructs of TPB. Therefore, all the hypotheses within the model were statistically supported in understanding determinants of mobile learning readiness. All in all, the study approved the effectiveness of well-structured cognitive psychological model in understanding pre-service teachers' intention towards adoption of mobile learning in the Turkish context. The study has important implications for researchers, educators, education stakeholders, policy makers and mobile learning application designers.

Keywords: Mobile learning readiness, Pre-service teachers, Theory of planned behavior

1. Introduction

The popularity of mobile devices is increasing and is being used in many areas of the education process (Zydney & Warner, 2016). Therefore, due to the increased affordability and functionality, mobile learning has been involved in a new study trend that attracts many researchers to discover this technology, examine its impact on students and academicians, and develop the necessary infrastructure (Al-Emran, Elsherif & Shaalan, 2016; Kinash, Brand, & Mathew, 2012). Mobile learning can be defined as "*the process of development, delivery, and consumption of learning material via a learning system subscriber using mobile devices*" (Yousafzai, Chang, Gani, & Noor, 2016, p. 785). Mobile learning is accomplished with the use of a wide range of mobile devices such as data-travelers, mp3 and mp4 players, digital cameras, tablets, laptops, netbooks, personal digital assistants, pads, pods, and smartphones (Pargman, Nouri & Milrad, 2018; Şad & Göktaş, 2014).

Considering the education process, a great potential positive effect of mobile technology on learning was accepted by researchers (e.g., Nikolopoulou, Gialamas, Lavidas, & Komis, 2020; Wang, Hwang, Yin, & Ma, 2020) and mobile learning has numerous benefits to achieve education goals. It promotes established and context-aware learning, facilitates personalized learning, forges a link between formal and informal learning, improves continuous learning, and develop collaboration and communication between learning societies as it was involved in UNESCO Policy Guidelines (West & Vosloo, 2013). Furió, Juan, Seguí, and Vivo (2015) stated that because of mobile technology, children have the chance to find out what they have learned from a diverse range of different points of view. In addition, mobile devices offer a number of different ways to promote authentic learning environments in the courses (Cotič, Plazar, Starčič, & Zuljan, 2020) and enable students to connect with courses (Sullivan, Slater, Phan, Tan, & Davis, 2019).

Although mobile technologies have a very important place about informing teachers (Kim & Kim, 2017), mobile learning is not adequately theorized, especially in teacher education (Sánchez-Prieto, Hernández-García, García-Peñalvo, Chaparro-Peláez, & Olmos-Migueláñez, 2019). There is a limited number of teacher development or training activity related to mobile learning (Carpenter, Krutka, & Trust, 2016). Therefore, teacher education programs are in need of providing theoretically and pedagogically effective mobile learning technologies due to several reasons such as the fast growth of mobile devices as an effective teaching tool in the teaching process and the pressure of providing teachers with effective technology integration skills (Newhouse, Williams, & Pearson, 2006). Especially, including mobile learning in teaching process of pre-service teachers (PST) has great importance on increasing quality in future generations' education (Baydas & Yilmaz, 2018; Sánchez-Prieto et al., 2019). Moreover, PST have an important role on mobile technology adoption, because their capabilities and views have the longtime effect and potential to affect change (Tondeur, van Braak, Ertmer, & Ottenbreit-Leftwich, 2017)

Accordingly, a better comprehension of the mobile technology adoption process in the education of PST will help scholars and stakeholders collaborate to implement proper methods and strategies for mobile learning. In this vein, understanding PST' mobile learning intention to adopt mobile learning provides important clues about how to include mobile learning in the classroom environment throughout their professional life. To determine the factors predicting mobile learning intentions, some valid and well-established models were preferred by researchers (e.g., Sidik & Syafar, 2020). Among them, Theory of Planned Behavior (TPB) developed by Ajzen (1991) has been mostly used as a theoretical framework. Therefore, the current study aims to predict PST' intentions towards mobile learning readiness (MLR) with the framework of TPB which is derived from self-interest motive and includes rational considerations. There are some empirical supports confirming the importance of psychological factors on MLR of PST implying that PST have the possibility of using technologies (e.g., Sánchez-Prieto et al., 2019; Teo, Zhou, & Noyes, 2016). However, no studies have been conducted to examine the role of pro-social roles in forming PST' behavioral intentions within the rational-choice model. Therefore, it is unclear whether the rational antecedents proposed by the TPB are also predictive of PST' MLR. In addition, it is yet unclear the extents to which PST have positive attitudes, subjective norms and, perceived behavioral control (PBC) and whether such psychological variables are related to their MLR. Moreover, it is not clear to what extent salient beliefs including behavioral beliefs, normative beliefs and control beliefs have an effect on the rational considerations. Therefore, to increase understanding of PST' mobile learning readiness, the study attempted to test a model that investigated the importance of salient beliefs on attitudes, subjective norms and, PBC which in turn influence intention to adopt mobile learning. Accordingly, it is expected that the study will make significant contributions to the literature in order to establish a solid foundation of the importance of mobile technologies in teacher education since it is recommended that teachers give more place mobile devices in their classrooms and professional development (Baran, 2014).

2. Literature review and hypothesis development

2.1. Mobile learning in pre-service teacher education

Mobile learning provides a number of benefits to teachers such as reflection of professional learning as a critical component of action (Kearney & Maher, 2019), younger generations such as PST who are more comfortable using technology are inclined to use mobile technology in their classrooms in future (Thomas & Bannon, 2013). However, although it could be emphasized that PST are in need of teaching models on how to integrate technology into the education process and should understand the pedagogical role of technology in education (Baydas & Yilmaz 2018), previous studies indicated that using a mobile learning environment can provide successful results in PST' education (Baran, 2014; Olpak & Ateş, 2018). Mobile technology has the makings of promoting PST to grasp and improve new consciousness (Husbye & Elsener, 2013), conduct a scientific study (Gado, Ferguson, & van't Hooft, 2006) and improve professional training (Kearney & Maher, 2019) and ease mentoring through collaboration (Husbye & Elsener, 2013).

2.2. Theory of planned behavior

TPB was developed from the earlier study of Theory of Reasoned Action (Ajzen & Fishbein, 1980). The theory explains systematically investigating the factors affecting behavioral choices (Ajzen, 1991). According to the TPB, the closest prediction of behavior is the intention which is individuals' readiness to implement a specific behavior and determined by attitude, subjective norm, and PBC (Ajzen & Madden, 1986). This section includes definitions of each factor and purposed hypotheses in compliance with the proposed model outlined in Figure 1.

2.2.1. Attitude

Attitude (ATT) can be defined as a "*learned predisposition to respond in a consistently favourable or unfavourable manner with respect to a given object*" (Fishbein & Ajzen, 1975, p. 6). The theory claims that in case people have more positive attitude towards a particular behavior, they are more chances to act that behavior (Ajzen, 1991). Mobile learning can increase collaboration between students and promote interaction among them and educators (Al-Emran et al., 2016). Therefore, attitudes towards educational technology can be used to gauge whether this technology has a positive or negative influence on the classroom environment (Ardies et al., 2015). However, the number of the studies investigating pre-service or in-service teachers' attitudes toward mobile learning adoption is limited (e.g., Kim & Kim 2017; Thomas & Bannon, 2013), within our knowledge, there is no empirical study examining the relationship between PST' attitudes toward mobile learning and their intention to adopt mobile learning. Based on earlier studies (e.g., Al-Emran et al., 2016), the study assumed that there is a positive relationship between PST' attitudes and their MLR. On the basis of this discussion, the following hypothesis can be proposed:

H1: Attitude toward mobile learning is positively related to intention (INT) to adopt mobile learning.

2.2.2. Subjective norm

Subjective norm (SN) reflects assessment and support of persons who are important to him/her such as family members, friends, co-workers for a behavior (Werner, 2004). In the educational technology context, although teachers are the most important decision-maker in the class, social pressure has an important effect on the decision (Teo, 2015). For this reason, teachers are conscious that some people such as colleagues and school principal have some expectations regarding the use of mobile devices for teaching in lessons (Sánchez-Prieto et al., 2019). Before teaching career, PST tend to receive feedback and value it more positively since they have some deficiencies in terms of professional development and experience (Lamote & Engels, 2010). Therefore, the study assumed that the perception that PST are pressured to use mobile technologies may influence their intention to adopt mobile learning. Accordingly, in a limited number of similar studies conducted with PST (e.g., Sánchez-Prieto et al., 2019), past empirical studies emphasized that the subjective norm is related to intention to adopt mobile learning. The discussion leads to the following hypothesis:

H2: Subjective norm is positively related to intention to adopt mobile learning.

2.2.3. Perceived behavioral control

PBC is referred to as an individual perceived ease or difficulty of performing a particular behavior and predicts behavioral intention (Ajzen, 1991). In the study context, PBC is defined as PST' perception of the ease or difficulty of adopting mobile learning. Therefore, we assumed that PST' intentions toward mobile learning depend on their confidence in mastering the new learning approach and their perceptions about acting a behavior as ease or difficulty have an influence on their readiness toward mobile learning. In a study conducted by Teo et al. (2016), a positive relationship was found between teachers' PBC and technology usage intentions. Therefore, the hypothesis is proposed as follows:
H3: PBC is positively related to intention to adopt mobile learning.

2.2.4. Salient beliefs

Behavior is a function of salient beliefs which are thought to be important predictors of people's intentions in the TPB, three salient beliefs were postulated (behavioral beliefs, normative beliefs, and control beliefs (Ajzen, 1991). Among them, behavioral beliefs are related to attitudes; normative beliefs are assumed to be the main predictors of subjective norms and control beliefs underlies perceptions toward behavioral control (Fishbein & Ajzen, 1975). Considering the mobile learning studies, salient beliefs were also involved in the framework of behavioral models or theories such as TAM (e.g., Huang, Teo, & Scherer, 2020) and TPB (e.g., Gómez-Ramirez et al., 2019). Looking at behavioral beliefs, perceived ease of use (PEOU) can be defined as "*the degree to which a person believes that using a particular system would be free of effort*" and perceived usefulness (PU) as "*the degree to which a person believes that using a particular system would enhance his or her job performance*" (Davis, 1989, p. 320). PU and PEOU were approved as an essential determinant of the technology acceptance and applicability of mobile technology systems (Al-Emran, Mezhuyev, & Kamaludin, 2020; Teo et al., 2016). Further, the concepts of PEOU and PU were also involved in many studies because of a significant impact on attitude to use mobile technology and mobile learning acceptance referring that people who see mobile technology as useful or easy to use are more likely to adopt it (e.g., Briz-Ponce Pereira et al., 2017). Earlier studies conducted with pre-service and in-service teachers proved that attitude toward the use of mobile technologies was influenced by PEOU and PU (e.g., Bower, DeWitt, & Lai, 2020; Teo et al., 2016). Therefore, the study proposes the following hypotheses:

H4: PEOU is positively related to attitude toward mobile learning.
H5: PU is positively related to attitude toward mobile learning.

Considering normative beliefs which are the second salient belief type, the subjective norm is influenced by existing normative beliefs that explain the anticipations of other people as an essential factor in intention (Ajzen, 1991). The scope of the study is interested in normative beliefs as perceptions towards to what extent other people are in favor of using mobile technology in their classroom. Earlier studies indicated that instructors and peer students are considered as prominent referent sets in higher education (Taylor & Todd, 1995), so they are thought to be the antecedents of subjective norm. Based on this discussion, the hypotheses are developed as follows:

H6: Instructor readiness (IR) is positively related to subjective norm for mobile learning.
H7: Student readiness (SR) is positively related to subjective norm for mobile learning.

Control beliefs, the last salient variable in the model, can be affected by the experience of friends, acquaintance, and other factors that enhance or reduce the perceived difficulty related to exhibiting the behaviors (Ajzen & Driver, 1991). Among the beliefs, perceived self-efficacy can be defined as "*people's beliefs about their capabilities to exercise control over their own level of functioning and over events that affect their lives*" (Bandura, 1991, p. 257). Individuals who believe they have an ability to act a certain skill or activity are inclined to intend to do the skill or perform the activity (Cheon et al., 2012). In line with the theoretical framework, if an individual has self-efficacy toward computers at high level, he/she is also inclined to higher levels toward the usage of information technology (Compeau & Higgins, 1995). In addition, the concept of learner autonomy involved in the scope of control beliefs is related to participation, control and evaluation and autonomous learners are responsible for all decisions regarding all directions of learning including specifying aims, describing the scope and the advancement, determining method, monitoring procedure related to acquiring the proper speech and evaluating acquired (Holec, 1981). Since mobile learning is regarded to be self-disciplined and self-motivated as necessary (Liu, 2008), learning autonomy could be a significant factor of behavioral control for mobile learning (Cheon et al., 2012). Hence, the hypotheses are formulated as:

H8: Perceived self-efficacy (SE) toward mobile learning is positively related to PBC.
H9: Learning autonomy (LA) toward mobile learning is positively related to PBC.

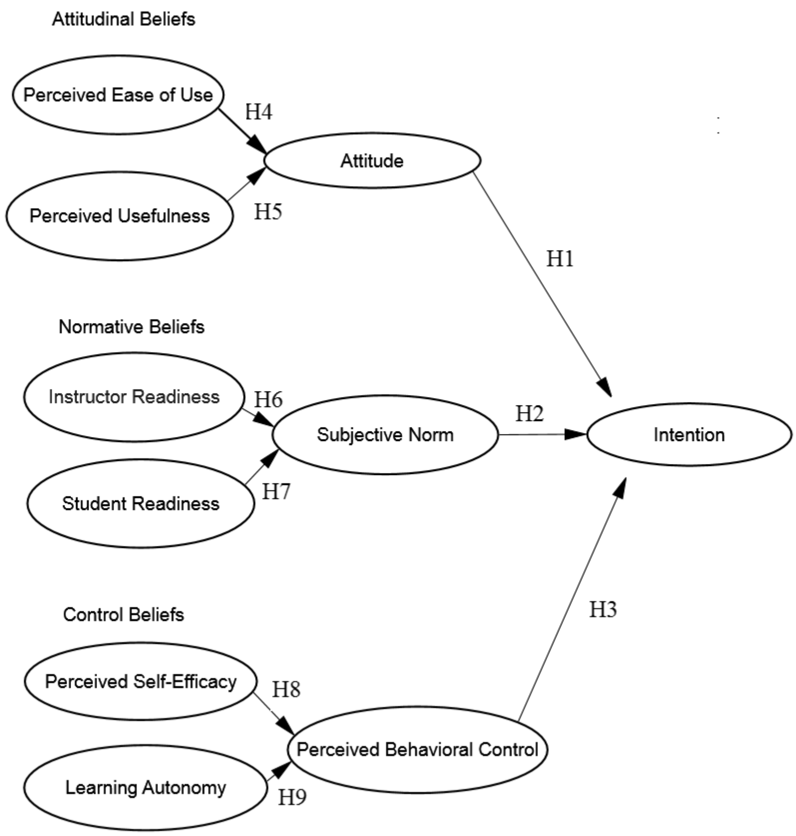


Figure 1. Research model of the study adapted from Cheon et al. (2012)

3. Methodology

3.1. The instrument

The Mobile Learning Readiness Scale (MLRS) was used as a data collection instrument. The original version of the MLRS was obtained from Cheon et al. (2012). In addition, the MLRS was adapted to the Turkish language by Sungur Gül and Ateş (2021). MLRS based on TPB consists of 30 items (three items for each of the 10 factors); 3 items for PEOU, 3 items for PU, 3 items for ATT, 3 items for IR, 3 items for SR, 3 items for SN, 3 items for SE, 3 items for LA, 3 items for BC and finally 3 items for INT to adopt mobile learning. All of the items on the scale are positively coded. The respondents of the scale are required to indicate on a Likert scale of totally disagree (1) to totally agree (7).

3.2. Sample and data collection

The participants of the study are 533 PST (62 males, 471 females) who were determined by using the convenience sampling method and participated voluntarily educating at two different faculties of education in the fall semester of 2019-2020 in Turkey. The reason for choosing these universities is that PST have been introduced to mobile applications in their content, content education and pedagogy courses and the universities' accessibility to mobile learning environments is similar. In addition, the students actively participated in the lessons by sharing files and voice using mobile applications via moodle in distance education. Hair, Babin, Anderson, and Black (2018) noted that the sample size was 10-15 sample/item for applying structural equation modeling (SEM). The current sample size of 533 with ten constructs of 30 items was also considered to be fit and above (533 > 30*15 = 450) the desired level. Therefore, the sample size was considered to be appropriate. The descriptive statistics are expressed in Table 1.

Table 1. Demographic characteristics

Characteristic	Demographic	Frequency	%
Gender	Male	62	11.63
	Female	471	88.37
Department	Preschool education	104	19.43
	Primary education	86	16.14
	Science education	343	64.43
Grade level	1st grade	157	29.45
	2nd grade	123	23.07
	3rd grade	148	27.78
	4th grade	105	19.70
Using mobile devices in education	Yes	448	84.05
	No	85	15.95
Using mobile devices in daily life	Yes	504	94.56
	No	29	5.44
The duration of mobile devices in daily use	Less than 1 hour	39	7.32
	1–4 hours	213	39.96
	5–8 hours	195	36.59
	More than 9 hours	86	16.14

The departments of the participants were science education, preschool education, and primary education in two mid-sized universities. In Turkey, students in the 3-6 years are educated by PST who graduated from the department of preschool education, students in the 7-10 years are educated by PST who graduated from the department of primary education, and students in the 11-14 years are educated by PST who graduated from the department of science. For preschool and primary school students, mobile applications such as sky map, anatomy 4D, Geogebra classic, geometry pad make learning concrete and provide fast access to information.

3.3. Data analysis

The proposed research model was analyzed using SPSS V.20 & AMOS V.24 software, respectively. Two staged SEM procedure suggested by Anderson and Gerbing (1988) was used. At the first stage, Confirmatory Factor Analysis (CFA) was performed and checked reliability and validity among items and constructs by using measurement model. In the second stage, the structural model was evaluated by testing the model fit, research hypothesis, and moderator effects.

4. Results

4.1. Measurement model: Reliability and validity

Prior to CFA, we performed Keiser-Meyer-Olkin (KMO) and Bartlett's test of sphericity to evaluate multivariate normality and sampling adequacy. Since Bartlett's test of sphericity showed a significant value ($p < 0.01$) and KMO value (0.92) was higher than 0.60 (Tabachnick & Fidell, 2019), the data was suitable for factor analysis. We also examined univariate statistics for all the study items to assess multivariate normality. (see Table 2). Tabachnick and Fidell (2019) suggested that univariate skewness values should be < |2| and univariate kurtosis values should be < |4| to provide a normal distribution. And then index of multivariate kurtosis and its critical ratio (c.r) were evaluated. Mardia's multivariate kurtosis estimate (Mardia, 1970) in the study was 408.30 lower than 960 calculated according to the formula, p (p + 2) (Raykov & Marcoulides 2008). (p: the number of observed variables in the model). Thus, the assumptions of multivariate normality were provided in this study.

Table 2. Descriptive statistics and measurement model: Reliability and validity

Construct	Items	Mean	SD	FL	CA	CR	AVE	Skewness	Kurtosis
PEOU	PEOU1	5.58	1.14	0.75	0.79	0.80	0.57	-1.26	1.75
	PEOU2	5.73	1.06	0.80				-1.52	3.46
	PEOU3	5.49	1.01	0.72				-0.71	0.53
PU	PU1	5.55	1.10	0.68	0.70	0.70	0.43	-0.95	1.23
	PU2	5.74	1.07	0.58				-1.18	1.88
	PU3	5.68	1.11	0.71				-1.29	2.41
ATT	ATT1	4.77	1.43	0.55	0.78	0.70	0.44	-0.47	-0.09
	ATT2	5.26	1.28	0.70				-1.00	1.25
	ATT3	5.43	1.20	0.72				-0.93	1.02
IR	IR1	5.28	1.24	0.73	0.61	0.63	0.38	-0.86	0.70
	IR2	5.21	1.28	0.67				-0.91	0.69
	IR3	4.85	1.40	0.39				-0.67	-0.06
SR	SR1	5.20	1.29	0.72	0.67	0.69	0.44	-0.80	0.37
	SR2	5.19	1.26	0.75				-0.78	0.48
	SR3	4.79	1.37	0.48				-0.41	-0.29
SN	SN1	5.21	1.26	0.72	0.79	0.76	0.51	-0.80	0.63
	SN2	5.32	1.21	0.70				-0.96	1.15
	SN3	5.42	1.18	0.73				-0.87	0.88
SE	SE1	5.35	1.26	0.76	0.77	0.77	0.52	-1.03	1.24
	SE2	5.40	1.27	0.69				-0.90	0.66
	SE3	5.30	1.23	0.72				-0.86	0.76
LA	LA1	5.66	1.06	0.75	0.79	0.79	0.56	-1.05	1.71
	LA2	5.65	1.07	0.80				-1.11	1.74
	LA3	5.49	1.15	0.70				-1.07	1.62
BC	BC1	5.18	1.38	0.57	0.82	0.66	0.39	-0.89	0.54
	BC2	5.26	1.31	0.64				-0.84	0.64
	BC3	5.54	1.25	0.66				-1.08	1.17
INT	INT1	5.44	1.21	0.78	0.78	0.76	0.52	-1.03	1.23
	INT2	5.54	1.11	0.74				-1.01	1.23
	INT3	5.26	1.37	0.64				-0.92	0.64

According to the two stage SEM procedures suggested by Anderson and Gerbing (1988), in the first stage, the information about convergent and divergent validity was accessed via CFA on the measurement model. Cronbach's Alpha (CA) and composite reliability (CR) test was performed to assess reliability. CA was applied to measure the internal consistency among items, while CR was utilized to describe the extent to which a train of items can represent potential constructs. In the current study, the factor loadings (FL) were higher than 0.30, Cronbach α values were ranged from 0.61 to 0.82, CR values were ranged from 0.63 to 0.80 and finally, the values of Average Validity Extracted (AVE) of variables were ranged from 0.38 to 0.57 as seen in Table 2. Fornell and Larcker (1981) suggested that if AVE is less than 0.50, but CR is higher than 0.60, the convergent validity of the construct is still satisfactory. In the literature, the minimum acceptable value for CA was suggested as 0.60 (Hair et al., 2018).

Lastly, since the values of the square root of AVE were all greater than the correlation between a pair of research constructs, sufficiency of discriminant validity seen in Table 3 was provided (Fornell & Lacker, 1981). Goodness of fit indices was examined with the following assessment criteria: χ^2 /df value should be between 2 and 5 (Hair et al., 2018), CFI and GFI values should be higher than 0.90 (Kline, 2005), IFI should be above 0.90 (Schumacher & Lomax, 2004), RMSEA and RMR should be less than 0.08 (Hu & Bentler, 1999). Therefore, it was determined that value of CFA fit indices represented a good model fit (Goodness-of-Fit Statistics: χ^2 = 772.97, df = 359, $p < 0.001$, χ^2 /df = 2.150, RMSEA = 0.047, CFI = 0.99, GFI = 0.91, IFI = 0.99, SRMR = 0.039).

Table 3. Correlation between the constructs, construct mean and standard deviation

	PEOU	PU	ATT	IR	SR	SN	SE	LA	BC	INT
PEOU	**0.757**									
PU	0.594	**0.659**								
ATT	0.458	0.563	**0.727**							
IR	0.412	0.438	0.402	**0.613**						
SR	0.452	0.397	0.417	0.492	**0.663**					
SN	0.455	0.481	0.421	0.465	0.583	**0.743**				
SE	0.477	0.525	0.546	0.441	0.492	0.544	**0.721**			
LA	0.551	0.567	0.487	0.430	0.464	0.509	0.640	**0.748**		
BC	0.444	0.360	0.347	0.337	0.517	0.422	0.563	0.530	**0.772**	
INT	0.478	0.508	0.516	0.396	0.492	0.578	0.653	0.632	0.653	**0.736**
M	5.60	5.65	5.15	5.11	5.06	5.32	5.35	5.60	5.33	5.41
SD	0.89	0.86	1.08	0.98	1.01	1.02	1.03	0.92	1.13	1.04

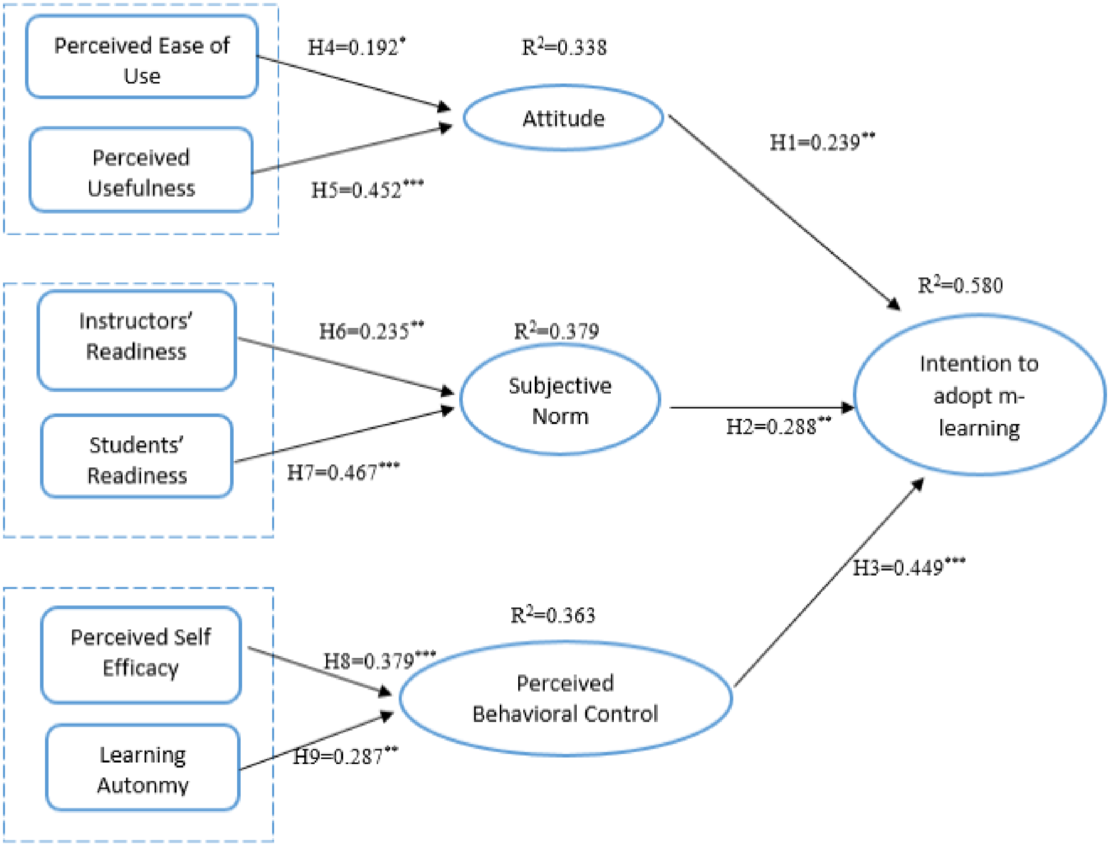
Note. Diagonal values show the square of AVE.

4.2. Structural model: Goodness of fit statistics, hypotheses testing, and modeling comparisons

SEM analysis revealed that goodness of fit statistics of theoretical framework represented a good fit (χ^2 = 877.04, df = 380, $p < 0.001$, χ^2 /df = 2.308, RMSEA = 0.050, CFI = 0.98, GFI = 0.90, IFI = 0.98, SRMR = 0.045). Assessment of the direct effects between the research constructs was performed, and the results were as follows: Attitude ($\beta = 0.239$, $t = 6.145$, $p < 0.01$), SN ($\beta = 0.288$, $t = 6.896$, $p < 0.01$) and PBC ($\beta = 0.449$, $t = 8.478$, $p < 0.001$) were significant in determining the INT to adopt mobile learning. PEOU ($\beta = 0.192$, $t = 4.368$, $p < 0.05$) and PU ($\beta = 0.452$, $t = 8.634$, $p < 0.001$) had a significant positive influence on attitude. IR ($\beta = 0.235$, $t = 5.995$, $p < 0.01$) and SR ($\beta = 0.467$, $t = 9.889$, $p < 0.001$) showed significant effects on SN. Lastly, SE ($\beta = 0.379$, $t = 7.421$, $p < 0.001$) and LA ($\beta = 0.287$, $t = 6.867$, $p < 0.01$) had a significant positive effect on PBC. In addition, the proposed model had superior ability for predicting ATT ($R^2 = 0.338$), SN ($R^2 = 0.379$), PBC ($R^2 = 0.363$) and INT ($R^2 = 0.580$). Thus, all hypothesis were supported and shown in Table 4 and Figure 2.

Table 4. Structural Equation Modeling results of the proposed model

Hypotheses	Standardized estimate	*t*-value	Hypothesis
H1: ATT → INT	0.239	6.145	Supported
H2: SN → INT	0.288	6.896	Supported
H3: PBC → INT	0.449	8.478	Supported
H4: PEOU → ATT	0.192	4.368	Supported
H5: PU → ATT	0.452	8.634	Supported
H6: IR → SN	0.235	5.995	Supported
H7: SR → SN	0.467	9.889	Supported
H8: SE → PBC	0.379	7.421	Supported
H9: LA → PBC	0.287	6.867	Supported



Note. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Figure 2. Result of Structural Equation Modelling

4.3. Moderating effects of personal characteristics

Invariance test for measurement and structural models program was conducted to explore the effect of PST' department and grade level as moderators in their mobile learning readiness. In the process, non-restricted model and full-metric invariance were produced. As seen in Table 5, the values of goodness of fit indices for models were favorable and according to the Chi-square difference test, two models were no significantly different. Thus, full metric invariance supported. Chi-square difference was compared between the baseline and the nested models according to the difference in the degrees of freedom (Byrne, 2016). The baseline model's goodness of fit indices was acceptable for department and grade level.

Table 5. Results for the moderating role of department

Measurement invariance model for grade levels				
Models	χ^2	df	$\Delta\chi^2$	Full metric invariance
Non-restricted model	331.430	62	$\Delta\chi^2(19)$= 34.30 p > 0.01 (insignificant)	Supported
Full-metric invariance model	297.127	43		
Other goodness of fit indices of non-restricted model: RMSEA=0.057; CFI=0.92; IFI=0.92				
Other goodness of fit indices of full-metric invariance model: RMSEA=0.058; CFI=0.91; IFI=0.91				

Paths	Science		Primary		Preschool		Baseline model	Nested model
	β	t-values	β	t-values	β	t-values		
H1: ATT → INT	0.255	7.057***	0.193	3.035**	0.173	2.833**	χ^2 (63) = 357.127	χ^2 (65) = 359.186[a]
H2: SN → INT	0.294	7.745***	0.192	3.025**	0.346	5.442***		χ^2 (65) = 359.320[b]
H3: PBC → INT	0.361	10.131***	0.676	10.537***	0.511	9.084***		χ^2 (65) = 365.122[c]
H4: PEOU →ATT	0.281	4.296***	0.110	1.173	0.136	1.147		χ^2 (65) = 358.536[d]
H5: PU → ATT	0.584	8.478***	0.625	6.691***	0.371	2.890**		χ^2 (65) = 363.297[e]
H6: IR → SN	0.218	4.013***	0.246	2.772**	0.324	4.066***		χ^2 (65) = 358.355[f]
H7: SR → SN	0.433	8.333***	0.546	6.150***	0.557	7.053***		χ^2 (65) = 358.946[g]
H8: SE → PBC	0.387	6.409***	0.466	4.682***	0.408	3.117**		χ^2 (65) = 358.321[h]
H9: LA → PBC	0.325	4.701***	0.304	3.054**	0.416	3.152**		χ^2 (65) = 357.607[i]

Goodness of fit indices of the baseline model for three groups: RMSEA = 0.052, CFI = 0.91, GFI = 0.91, IFI = 0.90.	[a] $\Delta\chi^2(2) = 2.059$ [b] $\Delta\chi^2(2) = 2.193$ [c] $\Delta\chi^2(2) = 7.995^{*}$ [d] $\Delta\chi^2(2) = 1.409$ [e] $\Delta\chi^2(2) = 6.170^{*}$ [f] $\Delta\chi^2(2) = 1.228$ [g] $\Delta\chi^2(2) = 1.819$ [h] $\Delta\chi^2(2) = 1.194$ [i] $\Delta\chi^2(2) = 0.480$

Note. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

The findings of multiple group analysis revealed that there were significant differences in the PBC-INT link and PU-ATT link. On the other hand, ATT-INT path, SN-INT path, PEOU-ATT path, IR-SN path, SR-SN path, SE-PBC path, LA-PBC path were not statistically significant differences between departments.

Table 6. Results for the moderating role of grade level

Measurement invariance model for grade levels				
Models	χ^2	df	$\Delta\ \chi^2$	Full metric invariance
Non-restricted model	215.375	82	$\Delta\ \chi^2(17)= 26.73$ p > 0.01 (insignificant)	Supported
Full-metric invariance model	188.642	65		
Other goodness of fit indices of non-restricted model: RMSEA=0.051; CFI=0.91; IFI=0.91				
Other goodness of fit indices of full-metric invariance model: RMSEA=0.052; CFI=0.90; IFI=0.90				

Paths	1st grade		2nd grade		3rd grade		4th grade		Baseline model	Nested model
	β	*t*-values	*β*	*t*-values	*β*	*t*-values	*β*	*t*-values		
H1: ATT → INT	0.272	2.994^{**}	0.110	1.447	0.184	2.276^{*}	0.103	1.414	χ^2 (84) = 216.642	χ^2 (87) = [a]218.549
H2: SN → INT	0.052	0.484	0.240	2.567^{*}	0.345	4.625^{***}	0.545	7.661^{***}		χ^2 (87) = [b]225.783
H3: PBC → INT	0.555	7.274^{***}	0.601	6.747^{***}	0.421	6.839^{***}	0.583	8.738^{***}		χ^2 (87) = [c]220.159
H4: PEOU → ATT	0.192	0.859	0.165	0.764	0.140	1.208	0.302	1.321		χ^2 (87) = [d]217.048
H5: PU → ATT	0.633	4.089^{***}	0.449	1.843	0.445	3.759^{***}	0.458	2.009^{*}		χ^2 (87) = [e]219.293
H6: IR → SN	0.379	3.311^{***}	0.091	0.681	0.350	3.677^{***}	0.276	2.165^{*}		χ^2 (87) = [f]219.802
H7: SR → SN	0.621	5.564^{***}	0.580	5.084^{***}	0.533	5.188^{***}	0.538	4.445^{***}		χ^2 (87) = [g]217.059
H8: SE → PBC	0.600	3.173^{**}	0.274	2.141^{*}	0.425	2.655^{**}	0.875	4.965^{***}		χ^2 (87) = [h]224.262
H9: LA → PBC	0.279	1.283	0.542	3.632^{***}	0.467	2.430^{*}	0.152	0.357		χ^2 (87) = [i]224.003
Goodness of fit indices of the baseline model for three groups: RMSEA = 0.048, CFI = 0.90, GFI = 0.91, IFI = 0.92.									[a] $\Delta\chi^2(3) = 1.907$ [b] $\Delta\chi^2(3) = 9.141^{*}$ [c] $\Delta\chi^2(3) = 3.517$ [d] $\Delta\chi^2(3) = 0.406$ [e] $\Delta\chi^2(3) = 2.651$ [f] $\Delta\chi^2(3) = 3.160$ [g] $\Delta\chi^2(3) = 0.418$ [h] $\Delta\chi^2(3) = 7.620$ [i] $\Delta\chi^2(3) = 7.361$	

Note. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Similarly, non-restricted model and full-metric invariance were produced for grade levels. As seen in Table 6, the values of goodness of fit indices for models were favorable and according to the Chi-square difference test, two models were no significantly different. Thus full metric invariance supported. Based on the findings from the moderator analysis, there were significant grade level differences on the SN-INT linkage. However, ATT-INT path, PBC-INT path, PEOU-ATT path, PU-ATT path, IR-SN path, SR-SN path, SE-PBC path, LA-PBC path were found not to be significantly different between grade levels.

5. Discussion and implications

Given today's world in which teaching-learning in the classroom environment is transforming into mobile learning tools very rapidly and the importance of mobile learning in the education of PST was stressed in past studies (e.g., Kearney & Maher, 2019). There is a need to investigate which factors predict PST' intentions towards mobile learning. Therefore, the current study aimed to examine this gap and investigate to reveal a deeper understanding of PST' intention to adopt mobile learning incorporating three critical salient beliefs — Attitudinal Beliefs, Normative Beliefs, and Control Beliefs –into the TPB model.

There was strong support for the proposed model and thus, results indicated that the model had a satisfactory fit to the data. The explanatory power of the study showed that the TPB model explained 58% of the variance in intention to adopt mobile learning. SEM results revealed that PBC was the most predictive antecedent for the intention to adopt mobile learning. Revealing this finding has an important role in the research area of MLR since PBC has been considered an effective determinant of people's intentions in higher education (Cheon et al., 2012). Considering the sample of the study, the finding implies that PST believe that they have controls and self-confidence over use of mobile learning and making decisions to adopt mobile learning. This is consistent with the idea of Ajzen (1985) who stated that PBC is increased in case people believe that they have the self-confidence that they can handle obstacles. These results supported most of the previous studies (e.g., Cheon et al., 2012; Raza et al., 2018). However, some researchers stated that the most effective variable on mobile adoption intention was subjective norm (e.g., Yeap et al., 2016), while others reported that the attitude was the most predictor variable on intention (e.g., Azizi & Khatony, 2019; Gómez-Ramirez et al., 2019). The most important reasons for this difference between the studies may be due to the fact the departments of the individuals studying at the university are different and they have various cultural characteristics.

Results of the study also suggested that attitude and subjective norm increase the individuals' intention toward the adoption of technology. Accordingly, it can be implied that PST think that they show higher concern for mobile learning in their coursework and the significant others' willingness to adopt mobile devices for learning is an important determinant for MLR. These findings were supported most of the past studies, while some studies found that subjective norm (e.g., Azizi & Khatony, 2019; Buabeng-Andoh, 2020) and PBC (e.g., Buabeng-Andoh, 2020) have no significant effect on intention to adopt mobile learning. Accordingly, considering the results of the current study, the first three hypotheses were supported.

Results of salient beliefs showed that PEOU and PU had a positive influence on attitude toward the use of mobile learning, explaining the variance by 34%. This finding implies that PST who believe that mobile learning is easy to use and had high feelings about its PU will have stronger attitude toward mobile learning. This result shows that H4 and H5 are supported. The findings are in line with earlier study results (e.g., Buabeng-Andoh, 2020) which revealed that university students who believe that mobile learning is easy and useful are more likely to take advantage of mobile learning tools for their classes.

The current study also revealed that IR and SR are positively associated with subjective norm, accounting for the variance by 38%. In other words, subjective norm is predicted by PST' accessible normative beliefs that explain other PST' expectations as a significant antecedent of intention to adopt mobile learning, therefore, H6 and H7 were supported. This finding is consistent with the earlier studies (e.g., Gómez-Ramirez et al., 2019). However, there were a few researches that revealed no significant effects of normative beliefs on subjective norm. For example, in a study conducted by Azizi and Khatony (2019), while IR had a significant influence, there was no significant influence of student readiness on the subjective norm. The different results obtained between the studies may be due to the different samples and the countries in which the study was conducted.

Further, the study revealed that SE and LA had positive influences on PBC. The control beliefs explained 36% of the variance in PBC. This finding which was consistent with earlier studies (e.g., Raza et al., 2018; Yeap et al., 2016) implies that PST have positive beliefs and self-confidence about their capabilities to use a mobile device for their course. It can be also implied that PST take high responsibility and gain control over the learning process in educational environments driven by mobile technology (Cheon et al., 2012). Accordingly, H8 and H9 were supported. Finally, results of the multi-group analysis indicated that the moderation effect of department and grade level on mobile learning readiness was significant. However, within our knowledge, none of the earlier studies examined the moderator influence of department and grade level on mobile learning readiness.

The study offers some theoretical and practical implications. The proposed model well explained PST' MLR. The present study confirmed that attitude, subjective norm, and PBC significantly establish a linkage with intention towards the adoption of mobile learning. The results suggest that PST' PEOU and PU significantly influence their attitude toward mobile learning. The result supports the idea of Davis (1993) who developed TAM, stating that both PEOU and PU play an essential role in explaining the attitude towards mobile learning. The results also suggest that PST' beliefs related to the readiness of instructors and readiness of students considerably influence subjective norm. The result expands the previous literature which pointed out that the readiness of individuals on mobile learning is important (Azizi & Khatony, 2019). Another theoretical implication relates to stating the important role of SE and LA on the PBC which in turn affected intention towards the adoption of mobile learning. This finding confirms the idea that people who think that they can be gifted in a particular skill or behavior and can control their behaviors are in the tendency to have stronger intentions to act the skill or act the behavior (Cheon et al., 2012).

The results present important practical implications to the researchers, educators, education stakeholders, policymakers, and mobile learning application designers. The study can help researchers to further examine the other constructs which can influence PST' intention towards the adoption of mobile learning. The study has also increased the education stakeholders and policy makers' understanding and knowledge about the determinants of MLR in the Turkish context. It has shown that Turkish PST cope relatively well with the inhibiting factor when using mobile devices in the classroom environment since PBC was found as the most significant antecedent of intention towards the adoption of mobile learning among constructs of the TPB. This emphasizes the importance of creating useful conditions with regards to availability that can facilitate and ease the individuals' use of mobile devices in lessons. For example, educators and education stakeholders can use different methods including providing adequate education and technical and managerial support to enhance PST' intentions toward mobile adoption of mobile learning (Hsia, 2016). In addition, the policymakers and mobile learning application designers should focus on the PST' attitudes towards mobile learning and subjective norm since they have important roles in influencing intention towards the adoption of mobile learning. Individuals' attitudes and subjective norm can be increased by making mobile devices easier to use in the classroom, which may shape a positive image of mobile learning among educators. This is an important implication for PST who will become teachers in the future. In order for mobile integration to be successful in PST, teacher educators must integrate the efficient use of mobile technology into their courses and establish connections with classroom practice (Husbye & Elsener, 2013).

6. Limitations and future studies

The present study has several limitations which should be considered for future studies. First of all, the demographic characteristics of the study are a group of PST within a university. As different results can be revealed using different sample groups, teacher-based study groups can't reflect the perspectives of different students in higher education. Therefore, it may be useful for future studies to be able to examine and compare different sample groups. There is a very serious disproportion in the distribution of some variables including using mobile devices in education and daily life and gender. This situation constitutes one of the most important limitations of this study. Therefore, it is not possible to perform statistical analysis and compare research results in terms of these variables. In addition, the data for this study were collected in Turkey and thus the fact that the findings obtained from this study cannot be generalized to other countries constitutes one of the limitations of this study. Cross-country comparison studies may be conducted for future studies to fully reflect the MLR. Further, the study only focused on PST. Therefore, it may not be concluded as a general representation of MLR in Turkey. The study can be repeated in students in other education levels such as elementary, secondary and high school or in-service teachers to reveal whether the findings are similar. Moreover, the present study is limited to determining intention rather than behavior. Thus, future researchers may use the actual behaviors regarding mobile learning implementation that can help to establish a connection between intentions towards the adoption of mobile learning and using mobile devices in education.

7. Conclusion

The TPB presents a very useful theoretical framework in understanding wide-ranging behavioral intentions across different research areas including the adoption of mobile learning. However, the present study is the first attempt in the Turkish context that has used TPB for predicting factors affecting PST' MLR. Results of the study confirmed the effectiveness of a well-structured cognitive psychological model in understanding PST' intention towards the adoption of mobile learning in the Turkish context. Moreover, the findings approved the feasibility of the TPB model and the inclusion of the salient beliefs enhanced the robustness and predictive power of the proposed model in measuring the MLR. Overall, it can be concluded that the salient beliefs are obtained to be having a significant positive impact on Turkish PST' attitudes, subjective norms, and PBC as well as on their intentions towards the adoption of mobile learning.

Acknowledgment

A part of this study was presented as an oral presentation in "International Pegem Conference on Education" in Diyarbakır, Turkey between 16^{th}- 19^{th} of September 2020.

References

Ajzen, I. (1991). The Theory of planned behavior. *Organizational behavior and human decision processes*, *50*(2), 179-211.

Ajzen, I., & Driver, B. L. (1991). Prediction of leisure participation from behavioral, normative, and control beliefs: An Application of the theory of planned behavior. *Leisure sciences*, *13*(3), 185-204.

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social beha*viour. Englewood Cliffs, NJ: Prentice-Hall.

Ajzen, I., & Madden, T. J. (1986). Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. *Journal of Experimental Social Psychology*, *22*(5), 453-474.

Al-Emran, M., Elsherif, H. M., & Shaalan, K. (2016). Investigating attitudes towards the use of mobile learning in higher education. *Computers in Human behavior*, *56*, 93-102.

Al-Emran, M., Mezhuyev, V., & Kamaludin, A. (2020). Towards a conceptual model for examining the impact of knowledge management factors on mobile learning acceptance. *Technology in Society*, 101247.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A Review and recommended two-step approach. *Psychological Bulletin*, 103, 411–423.

Ardies, J., De Maeyer, S., Gijbels, D., & van Keulen, H. (2015). Students attitudes towards technology. *International Journal of Technology and Design Education*, *25*(1), 43-65.

Azizi, S. M., & Khatony, A. (2019). Investigating factors affecting on medical sciences students' intention to adopt mobile learning. *BMC medical education*, *19*(1), 1-10.

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational behavior and human decision processes*, *50*(2), 248-287.

Baran, E. (2014). A Review of research on mobile learning in teacher education. *Educational Technology & Society*, *17*(4), 17-32.

Baydas, O., & Yilmaz, R. M. (2018). Pre-service teachers' intention to adopt mobile learning: A Motivational model. *British Journal of Educational Technology*, *49*(1), 137-152.

Bower, M., DeWitt, D., & Lai, J. W. (2020). Reasons associated with preservice teachers' intention to use immersive virtual reality in education. *British Journal of Educational Technology*, *51*(6), 2214-2232.

Briz-Ponce, L., Pereira, A., Carvalho, L., Juanes-Méndez, J. A., & García-Peñalvo, F. J. (2017). Learning with mobile technologies–Students' behavior. *Computers in Human Behavior*, *72*, 612-620.

Buabeng-Andoh, C. (2020). Exploring University students' intention to use mobile learning: A Research model approach. *Education and information technologies*, 1-16.

Byrne, B. M. (2016). *Structural equation modeling with AMOS (3rd Ed.).* New York, NY: Routledge.

Carpenter, J., Krutka, D., & Trust, T. (2016). PK-12 teachers' conceptualizations of professional learning networks. In G. Chamblee & L. Langub (Eds.), *Proceedings of SITE International Conference* (pp. 1936-1942). Savannah, GA: AACE.

Cheon, J., Lee, S., Crooks, S. M., & Song, J. (2012). An Investigation of mobile learning readiness in higher education based on the theory of planned behavior. *Computers & Education*, 59(3), 1054-1064.

Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly*, *19*(2), 189-211.

Cotič, N., Plazar, J., Starčič, A. I., & Zuljan, D. (2020). The Effect of outdoor lessons in natural sciences on students' knowledge, through tablets and experiential learning. *Journal of Baltic Science Education, 19*(5), 747-763.

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

Davis, F. D. (1993). User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *International Journal of Man-Machine Studies*, *38*(3), 475-487.

Fishbein, M., & Ajzen, I. (1975). *Intention and Behavior: An Introduction to theory and research*. Reading, MA: Addison-Wesley.

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

Furió, D., Juan, M. C., Seguí, I., & Vivó, R. (2015). Mobile learning vs. traditional classroom lessons: a comparative study. *Journal of Computer Assisted Learning*, *31*(3), 189-201.

Gado, I., Ferguson, R., & van't Hooft, M. A. R. K. (2006). Using handheld-computers and probeware in a science methods course: Preservice teachers' attitudes and self-efficacy. *Journal of Technology and Teacher Education*, *14*(3), 501.

Gómez-Ramirez, I., Valencia-Arias, A., & Duque, L. (2019). Approach to m-learning acceptance among university students: an integrated model of TPB and TAM. *International Review of Research in Open and Distributed Learning*, *20*(3). doi:10.19173/irrodl.v20i4.4061

Hair, J. F., Babin, B. J., Anderson, R. E., & Black, W. C. (2018). *Multivariate data analysis* (8th ed.). Hampshire, United Kingdom: Cengage Learning, EMEA

Holec, H. (1981). Learner autonomy and language learning. *Council of Europe*, *1981*, 66-74.

Hsia, J. W. (2016). The Effects of locus of control on university students' mobile learning adoption. *Journal of Computing in Higher Education*, *28*(1), 1-17.

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

Huang, F., Teo, T., & Scherer, R. (2020). Investigating the antecedents of university students' perceived ease of using the Internet for learning. *Interactive Learning Environments*, 1-17. doi:10.1080/10494820.2019.1710540

Husbye, N. E., & Elsener, A. A. (2013). To move forward, we must be mobile: Practical uses of mobile technology in literacy education courses. *Journal of Digital Learning in Teacher Education*, *30*(2), 46-51.

Kearney, M., & Maher, D. (2019). Mobile learning in pre-service teacher education: Examining the use of professional learning networks. *Australasian Journal of Educational Technology*, *35*(1), 135-148.

Kim, H. J., & Kim, H. (2017). Investigating teachers' pedagogical experiences with tablet integration in Korean rural schools. *Asia-Pacific Education Researcher, 26*(1–2), 107–116.

Kinash, S., Brand, J., & Mathew, T. (2012). Challenging mobile learning discourse through research: Student perceptions of Blackboard Mobile Learn and iPads. *Australasian Journal of Educational Technology*, *28*(4), 639-655.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Press.

Lamote, C., & Engels, N. (2010). The Development of student teachers' professional identity. *European Journal of Teacher Education*, *33*(1), 3-18.

Liu, Y. (2008, July). *An Adoption model for mobile learning*. Paper presented at the IADIS e-commerce 2008 conference, Amsterdam, The Netherlands.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*, 519–530.

Newhouse, C. P., Williams, P. J., & Pearson, J. (2006). Supporting mobile education for pre-service teachers. *Australasian Journal of Educational Technology*, *22*(3), 289-311.

Nikolopoulou, K., Gialamas, V., Lavidas, K., & Komis, V. (2020). Teachers' readiness to adopt mobile learning in classrooms: A Study in Greece. *Technology, Knowledge and Learning*, 1-25.

Olpak, Y. Z., & Ates, H. (2018). Pre-Service science teachers' perceptions toward additional instructional strategies in biology laboratory applications: Blended learning. *Science Education International*, *29*(2), 88-95.

Pargman, T. C., Nouri, J., & Milrad, M. (2018). Taking an instrumental genesis lens: New insights into collaborative mobile learning. *British Journal of Educational Technology*, *49*(2), 219-234.

Raykov, T., & Marcoulides, G. A. (2008). *An Introduction to applied multivariate analysis*. New York, NY: Routledge.

Raza, S. A., Umer, A., Qazi, W., & Makhdoom, M. (2018). The Effects of attitudinal, normative, and control beliefs on m-learning adoption among the students of higher education in Pakistan. *Journal of Educational Computing Research*, *56*(4), 563-588.

Şad, S. N., & Göktaş, Ö. (2014). Preservice teachers' perceptions about using mobile phones and laptops in education as mobile learning tools. *British Journal of Educational Technology*, *45*(4), 606-618.

Sánchez-Prieto, J. C., Hernández-García, Á., García-Peñalvo, F. J., Chaparro-Peláez, J., & Olmos-Migueláñez, S. (2019). Break the walls! Second-Order barriers and the acceptance of mLearning by first-year pre-service teachers. *Computers in Human Behavior*, *95*, 158-167.

Schumacher, R., & Lomax, R. (2004). *A Beginner's guide to structual equation modelling*. London, UK: Lawrence Erlbaum Assocıates, Publishers.

Sidik, D., & Syafar, F. (2020). Exploring the factors influencing student's intention to use mobile learning in Indonesia higher education. *Education and Information Technologies*, *25*(6), 4781-4796.

Sullivan, T., Slater, B., Phan, J., Tan, A., & Davis, J. (2019). M-learning: Exploring mobile technologies for secondary and primary school science inquiry. *Teaching Science*, *65*(1), 13-16.

Sungur Gül, K. & Ateş, H. (2021). Turkish Adaptation of mobile learning readiness scale in the framework of theory of planned behavior. *International Online Journal of Education and Teaching,8(2)*, 814-831.

Tabachnick, B., G. & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). New York, NY: Pearson.

Taylor, S., & Todd, P. A. (1995). Understanding information technology usage: A Test of competing models. *Information Systems Research*, *6*(2), 144-176.

Teo, T. (2015). Comparing pre-service and in-service teachers' acceptance of technology: Assessment of measurement invariance and latent mean differences. *Computers & Education*, *83*, 22-31.

Teo, T., Zhou, M., & Noyes, J. (2016). Teachers and technology: Development of an extended theory of planned behavior. *Educational Technology Research and Development*, *64*(6), 1033-1052.

Thomas, K., & O'Bannon, B. (2013). Cell phones in the classroom: Preservice teachers' perceptions. *Journal of Digital Learning in Teacher Education*, *30*(1), 11-20.

Tondeur, J., Van Braak, J., Ertmer, P. A., & Ottenbreit-Leftwich, A. (2017). Understanding the relationship between teachers' pedagogical beliefs and technology use in education: A Systematic review of qualitative evidence. *Educational Technology Research and Development*, *65*(3), 555-575.

Wang, Z., Hwang, G. J., Yin, Z., & Ma, Y. (2020). A Contribution-oriented self-directed mobile learning ecology approach to improving EFL students' vocabulary retention and second language motivation. *Educational Technology & Society, 23*(1), 16-29.

Werner, P. (2004). *Reasoned action and planned behavior*. In S. J. Peterson, & T. S.Bredow (Eds.). *Middle range theories: Application to nursing research* (pp. 125–147). Philadelphia, PA: Lippincott Williams & Wilkins.

West, M., & Vosloo, S. (2013). *Directrices de la UNESCO para las políticas de aprendizaje móvil* [UNESCO Guidelines for Mobile Learning Policies]. Paris, France: UNESCO.

Yeap, J. A., Ramayah, T., & Soto-Acosta, P. (2016). Factors propelling the adoption of m-learning among students in higher education. *Electronic Markets*, *26*(4), 323-338.

Yousafzai, A., Chang, V., Gani, A., & Noor, R. M. (2016). Multimedia augmented m-learning: issues, trends and open challenges. *International Journal of Information Management, 36*, 784–792. doi:10.1016/j.ijinfomgt.2016.05.010

Zydney, J. M., & Warner, Z. (2016). Mobile apps for science learning: Review of research. *Computers & Education*, *94*, 1-17.

Zhang, S., Gao, Q., Wen, Y., Li, M., & Wang, Q. (2021). Automatically Detecting Cognitive Engagement beyond Behavioral Indicators: A Case of Online Professional Learning Community. *Educational Technology & Society*, *24* (2), 58–72.

Automatically Detecting Cognitive Engagement beyond Behavioral Indicators: A Case of Online Professional Learning Community

Si Zhang[1*], Qianqian Gao[1], Yun Wen[2*], Mengsiying Li[3] and Qiyun Wang[2]
[1]Hubei Research Center for Educational Informationization, Faculty of Artificial Intelligence in Education, Central China Normal University, China // [2]National Institute of Education, Nanyang Technological University, Singapore // [3]School of Management, Wuhan College, China // djzhangsi@mail.ccnu.edu.cn // 2956954801@qq.com // yun.wen@nie.edu.sg // 1061224142@qq.com // qiyun.wang@nie.edu.sg
[*]Corresponding author

(Submitted August 16, 2020; Revised September 6, 2020; Accepted February 1, 2021)

ABSTRACT: Online discourse is widely used in diverse contexts of learning and professional training, but superficial interactions and digression often occur. In the face of these problems and the large-scale unstructured text data, the traditional way of learning analytics has been challenged in terms of providing timely intervention and feedback. In this paper, a workflow for automatically detecting in-service teachers' cognitive engagement in an online professional learning community is described. Discourse data of 1834 in-service teachers involved in a teacher professional development program was collected and processed using the Word2vec toolkit to generate lexical vectors. The method of vector space projection was used to calculate the new information contained in each post, cosine similarity was used to calculate topic relevance, and cluster analysis was used to explore in-service teachers' discourse characteristics. Results showed that in-service teachers' average contribution was 4.59 posts and the average length of each post was 39.47 characters in Chinese. In the mathematics online professional learning community, the average amount of new information contained in each post was 0.221 and in-service teachers' posts contained much new information in the early stages of online discourse. Most in-service teachers' posts were relevant to the discussion topic. Cluster analysis showed three different groups of posts with unique characteristics: high topic relevance with much new information, high topic relevance with little new information, and low topic relevance with little new information. Finally, limitations are discussed and future research directions are proposed.

Keywords: Learning communities, Computer-mediated communication, Evaluation methodologies, Interactive learning environments

1. Introduction

The advantages of the online professional learning community (OPLC) have been identified by previous researchers as being essential to support reflection and discussion that can improve in-service teachers' learning engagement and leverage changes in their classrooms (e.g., Xing & Gao, 2018). Discourse in online professional learning communities helps in-service teachers share teaching experiences, solve teaching problems together, and deepen understanding of the relationship between theory and practice (Tsai, 2012). Millions of in-service teachers participate in online professional development programs each year, and they spend a significant amount of time learning in OPLCs (e.g., Zhang, Liu, Chen, Wang, & Huang, 2017; Xing & Gao, 2018). Nevertheless, it would be too simple to assume that the establishment of an OPLC will automatically promotes in-service teachers' effective interaction and collaborative knowledge construction (Baran & Cagiltay, 2010). Superficial interactions and off-task messages often occur in in-service teachers' online discussions (e.g., Lantz-Andersson, Lundin, & Selwyn, 2018). Similarly, some in-service teachers do not contribute any new information, and this adversely affects the participation and sharing willingness of other members of the community (e.g., Zhang et al., 2017). Discourse in an OPLC is a social and cognitive interwoven process and the effectiveness of online discourse depends on both the iterative interaction and the active information exchange among community members (e.g., Chen, Fan, & Tsai, 2014). Meanwhile, content-related (or on-topic) discussions are more conducive to learning than off-topic discussions (e.g., Cho, 2016). According to the Community of Inquiry (CoI) framework, engaging learners in critical reflection and collaborative discussion fosters their high-order learning (Heilporn & Lakhal, 2020). The CoI framework defines three critical elements of online learning: cognitive presence, social presence, and teaching presence. Among the elements, cognitive presence is a central component in the model, which describes stages of inquiry-based learning, including problem presentation, knowledge synthesis, and evaluation (Garrison, Anderson, & Archer, 1999). Thus, the effective measures of teachers' cognitive engagement in the OPLC that support a movement toward understanding their cognitive presence and providing timely intervention and regulation are required.

Cognitive engagement refers to learners' psychological involvement in academic tasks (Fredricks, Blumenfeld, & Paris, 2004). Itis positively correlated with learners' academic performance (Wang, Wen, & Rosé, 2016). In terms of measuring cognitive engagement, the commonly used indicators include the number of posts, speed of response, or time on task (e.g., Kim, Park, Yoon, & Jo, 2016; Xie, Heddy, & Greene, 2019). It is often not enough to measure learners' cognitive engagement only by observable indicators, but combining quantitative dada and qualitative content analysis will provide more reliable results (e.g., Atapattu, Thilakaratne, Vivian, & Falkner, 2019). However, the analysis of the large-scale online discourse data (mainly textual data) raises methodological problems, including data collection and coding (Vieira, Parsons, & Byrd, 2018). Facing the sheer data volumes and the diversity of in-service teachers' language expression, the traditional way of data coding used in class discourse analysis, however, cannot address this dilemma.

In this study, we present an automatic discourse analysis approach using a language modeling technique called neural word embeddings (Word2vec) (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) to detect learners' active participation in discussions in the online professional learning community.

2. Literature review

2.1. Online professional learning community and online discourse

Many online professional learning communities have been built using social media tools (e.g., Twitter, Blog, WeChat), online learning platforms (e.g., Blackboard, Moodle), or course management systems (CMS) (e.g., Lantz-Andersson et al., 2018; Prestridge, 2019). Online professional learning communities provide learning resources for in-service teachers including video cases, online discourse, as well as reflections (e.g., Lantz-Andersson et al., 2018). Online discussion with experts or peers is useful for in-service teachers' professional development. In-service teachers have opportunities to think about teaching problems from multiple perspectives (Lee & Brett, 2015; Polizzi, Head, Barrett-Williams, Ellis, Roehrig, & Rushton, 2018). Moreover, online discussion improves the ability of teacher groups by solving teaching problems through role-playing and collaboration (Yang, 2016). In addition, online discussion helps to improve the interactions between the members in the online learning communities through questioning, interpreting, and elaborating (Cho, 2016).

Although online discourse is available in most online professional learning communities, it is primarily used as a supporting discussion between participants rather than for meaningful knowledge sharing and co-construction (e.g., Atapattu et al., 2019; Zhang et al., 2017). Prestridge (2019) argued that the critical discourse that helps to facilitate teachers' knowledge construction and transform classroom practice was only sustained for a small number of postings by teachers in a given thread. According to Gunawardena, Lowe, and Anderson (1997), social knowledge construction among learners in a constructivist learning environment needs to go through some key stages, such as restating the participant's position, negotiation or clarification of the meaning of terms, and negotiation of new statements. In this regard, effective discourse depends on negotiating existing information (given information) and constantly coming up with new information. However, repeated discussion often occurs, and some teachers enact the role of information consumer and information networker in online discussion (e.g., Prestridge, 2019; Tsiotakis & Jimoyiannis, 2016). In addition, commitment and co-construction of ideas in an online professional learning community can only be achieved when teachers are deeply engaged with discussion topics because the topic is authentic and requires active cognitive engagement and in-depth understanding of the problem (Xing & Gao, 2018). Hence, in-service teachers' online discourse data in the OPLCs provides a wealth of valuable information for educational researchers and practitioners to understand the learning process. This understanding is important because it can help teacher training managers make correct intervention policies, improve the quality of teacher training programs, and thus enhance in-service teachers' teaching abilities. However, the large-scale online discourse data in OPLCs raises methodological problems, including data coding and analysis. Facing the large-scale data volumes, the diversity of teachers' language expression, and the complexity of online discourse process, the traditional way of data collection and analysis in the general classroom, cannot deal with this dilemma.

2.2. Discourse analysis based on machine learning technologies

Discourse analysis was originally proposed in the linguistics field and has been widely used in the fields of sociology, education, and psychology (Harris, 1952). In education, especially in online learning communities, text is the most commonly used data for discourse analysis (Wu, Yu, & Wang, 2018). With the growing scale of text data, discourse analysis has become a key application area of learning analytics and educational data mining.

Discourse analysis has been used to reveal the process of interaction and knowledge construction, investigate learner's behaviors and discourse content, and acquire deeper understanding of engagement in collaborative learning (e.g., Peng & Xu, 2020).

There have been two main lines of research evaluating online discourse by using the machine learning method. Some studies analyze and visualize the learners' cognitive and social presences in online learning settings (e.g., Xing & Gao, 2018). Others investigate the semantic content and processes of learner's posts in understanding interactive patterns (e.g., GaŠević, Joksimović, Eagan, & Shaffer, 2019).

The identification of new versus given information and topic relevance within a text has been frequently studied by researchers of discourse analysis. The most frequently used method to measure similarity between two texts is keyword matching, including word matching, keyword matching, andweighted keyword matching. The keyword matching method operates by finding exact matches between two texts and some disadvantages of this method include the lack of emphasis on differential importance as to how much information a particular word may carry. The strength of Latent Semantic Analysis (LSA) to measure text similarity has been demonstrated in many research studies (e.g., Chen, Chen, & Sun, 2010; Sung, Liao, Chang, Chen, & Chang, 2016). Sung et al. (2016) used 262 Chinese articles to construct a latent semantic space with 250 dimensions and the LSA was used to compute the semantic similarity between pairs of sentences, such as students' summary, expert's summary, or the source text. On account of the concept of new versus given information is related but distinct from the concept of text similarity, Hu, Cai, Louwerse, Olney, Penumatsa, and Graesser (2003) adapted the standard LSA and proposed the LSA based measure called a span method to detect given and new information in written discourse.

Word2vec is a basic word embedding technology in the field of Natural Language Processing (NLP). Word embedding originates from vector space models and aims to quantify and categorize semantic similarities between words based on their distributional properties in large-scale of text data (Li, Li, Fu, Masud, & Huang, 2016). Word2vec has been used in web search (Bing, Niu, Li, Lam, & Wang, 2018), comment sentiment classification (Zhang, Xu, Su, & Xu, 2015), and text classification (Sinoara, Camacho-Collados, Rossi, Navigli, & Rezende, 2019). Compared with other data mining methods, the accuracy of Word2vec in text classification is more than 70% (Cerisara, Král, & Lenc, 2018). The Word2vec technique has been successfully applied to detect teachers' cognitive engagement (Atapattu et al., 2019), and is at the forefront of the research field (Hashimoto, Alvarez-Melis, & Jaakkola, 2015).

3. Conceptual framework

3.1. Given and new information

Identifying given and new information in a discourse has long been regarded as an important research issue. Many researchers define given and new information from different perspectives. According to Halliday (1967), given information refers to what can be recovered anaphorically or situationally from the preceding discourse, and new information, on the contrary, is not recoverable. Chafe (1976) defines given information as "knowledge which the speaker assumes to be in the consciousness of the addressee at the time of the utterance" (p. 30), and new information as "what the speaker assumes he is introducing into the addressee's consciousness by what he says" (p. 30). Based on integrating previous theoretical studies, Prince (1981) develops a taxonomy that can be used to hand-code discourse text for identifying given and new information. According to Prince (1981), there are three different levels of givenness. On the first level, givenness represents the sense of predictability or recoverability which is based on the assumes that "the hearer can predict or could have predicted that a particular linguistic item will or would occur in a particular position within a sentence" (p. 226). On the second level, givenness represents the sense of saliency which is based on the assumes that "the hearer has or could appropriately have some particular thing or entity in his or her consciousness at the time of hearing the utterance" (p. 228). On the third level, givenness represents the sense of shared knowledge which is based on the assumes that "the hearer knows or can infer a particular thing" (p. 230). In addition, some other concepts, such as the theme and rhyme, the primacy and recency, are also related to what can be defined as given or new information in a discourse text. In this study, however, the distinction of given and new information is operationally defined purely in terms of semantic recoverability. The following examples (originally from Prince (1981)) illustrates how given and new information can be judged based on the criterion of recoverability.

Example1. We got some beer out of the trunk.

Example2. We got some picnic supplies out of the trunk.

In the sense of recoverability, picnic supplies in Example2 would be given information, whereas beer in Example1 would be new information, although, in some sense, they look a little bit different.

3.2. Language modeling

Taking a large corpus of text as input, Word2vec produces a vector space in which words that share common contexts in the corpus will be located in close proximity to one another. Taking the Skip-Gram approach in Word2vec as an example, the generation process of word vectors is shown in Figure 1. The goal of the Word2vec process is to learn similar word vectors for two words which have similar contexts. At last, the word vector corresponding to each word is determined.

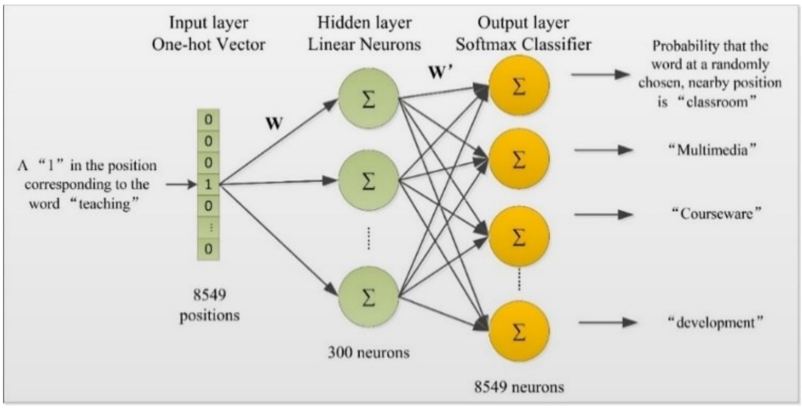


Figure 1. The generation process of word vectors (Skip-Gram approach)

4. Methodology

This study aimed to investigate the cognitive engagement, participation and discourse characteristics of the teachers in an online professional learning community. To achieve this goal, we formulated the first research question to understand teachers' nature of participation.
RQ1: What were the in-service teachers' participation characteristics in online discourse?

The second and third research questions focus on exploring teachers' cognitive engagement in the online professional learning community.
RQ2: How much new information was contained in each of the in-service teacher's post in online discourse?
RQ3: How relevant were the in-service teacher's posts to the discussion topic in online discourse?

The results would lead us to explore in-service teachers' online discourse characteristics through cluster analysis and answer the fourth research question.
RQ4. What were the in-service teachers' online discourse characteristics in terms of cognitive aspect?

4.1. Participants

In 2013, the Ministry of Education of China launched a five-year in-service teacher training program called the Information Technology Application Ability Enhancement Project for primary and secondary school teachers. They participated in the program in batches and the length of each training was about 120 hours. Many online learning platforms have been established and online learning was used in this project. In-service teachers needed to complete three tasks: watching video cases online, participating in topic-based online discussion, and submitting reflective diaries. The video cases were about how to use information technology tools to support classroom teaching. After watching the video cases, the in-service teachers in each community online discussed the contents of a video case based on their own teaching experience. In-service teachers participated in online discussion by posting, and each post is a reflection of the application of information technology in classroom teaching.

A total of 1,834 primary and secondary in-service teachers from a province in eastern China participated in the teacher professional development program for about a year. They were assigned to 31 online professional learning communities according to their subjects, with an average of 59.2 people per community. Table 1 shows the basic information of the 31 communities. After the program was completed, all online discourse data were collected and analyzed.

Table 1. Basic information of the 31 communities

Subject	Number of communities	Number of people	Subject	Number of communities	Number of people
Chinese	4	339	Mathematics	3	192
English	2	161	Geography	1	50
Chemistry	1	49	History	1	48
Art	2	48	Music	1	52
Biology	1	38	Ethics	2	95
Sports	2	111	Physics	2	54
Information technology	2	39	STEAM	3	266
Mixed group	1	42	Early childhood education	3	250

The purposive sampling method was used in this study to select one of the 31 communities to further analyze teachers' characteristics of cognitive engagement (Denzin & Lincoln, 1994). In the end, a mathematics online professional learning community was selected because it was representative (close to the average level of the population) in terms of gender, years of service and participation. The number of in-service teachers in the community was 60, and their average years of service was 19.3.

4.2. Discussion activities

In-service teachers' learning activities, such as watching video cases, participating in topic-based online discussion and submitting reflective diaries, were all carried out on the online learning platform. The interface of the online discussion activities is shown in Figure 2. The timeline of online discussion activities was as follows: (1) a teacher leader entered the name, duration, type and description of the topic-based discussion activity, and uploaded a video case, (2) all in-service teachers in the community watched the video case first and then participated in online discussions. All in-service teachers could participate in the discussion at any time. The basic learning requirement for in-service teachers was to contribute at least 5 posts during the online discourse session.

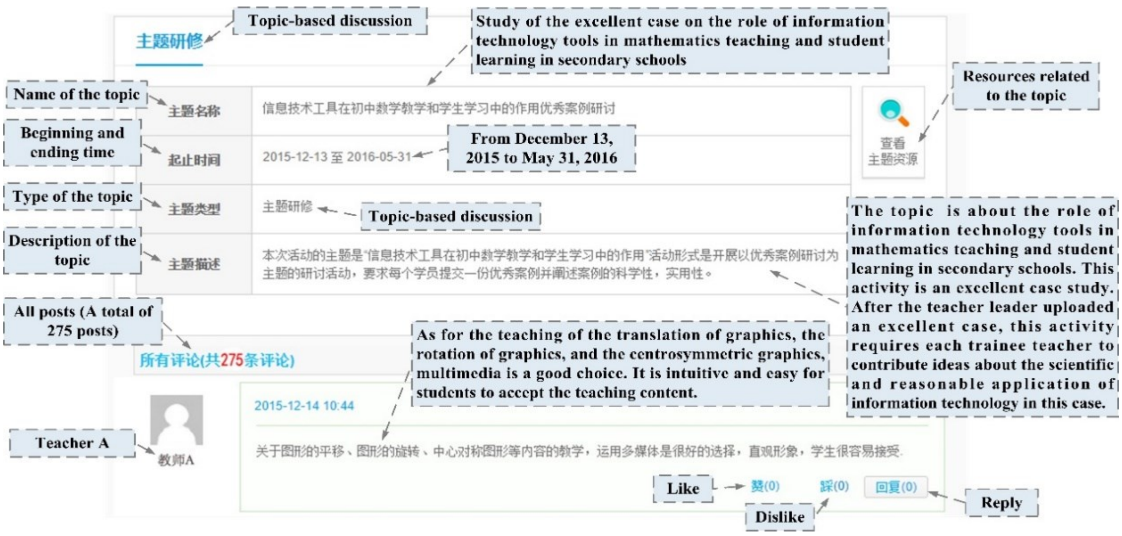


Figure 2. A screenshot of the online discussion interface

4.3. Data collection and analysis

After the teacher professional development program was completed, researchers used a web crawler tool called the Octopus to collect discourse data in all the 31 online professional learning communities. The research process consists of seven phases, as shown in Figure 3.

Phase 1: Data cleansing. All the collected data, including the name of the person who posted, the time and the text of the posts, were sorted out and stored in Excel files. All the in-service teachers' names were pseudonyms. The size of the Excel files is 3.09 M bytes. The number of posts of the data set "Application of information technology in classroom teaching" was 8418, and the average number of characters in Chinese per post was 39.47. The size of the set of lexical items in Chinese was 3859.

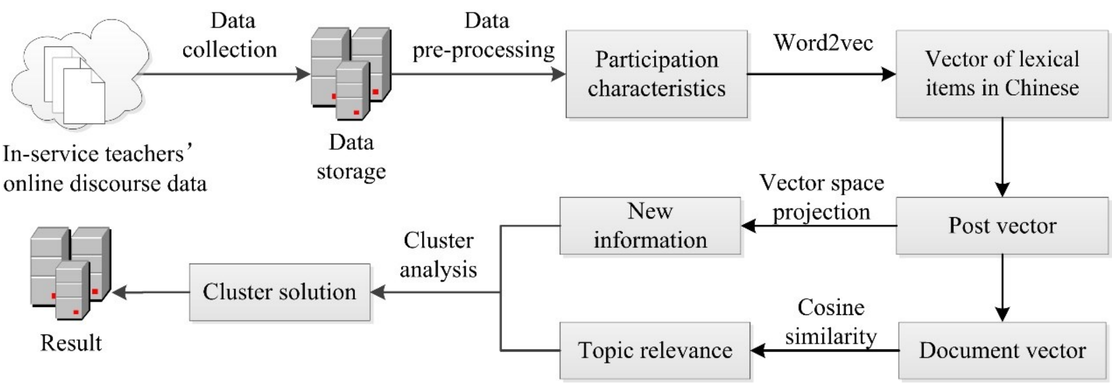


Figure 3. The process of data collection and analyses

Note. The black arrow represents the process of data analysis, and the rectangular box represents the result of data analysis.

Phase 2: Data preprocessing. The first author of this paper preprocessed the text data for all posts using the jiebaR package in R language, including Chinese segmentation, removing stop words, numbers, emojis, and special symbols. In particular, we used feature engineering methods (e.g., TFIDF, Singular Value Decomposition) to select professional terms (e.g., classroom teaching, multimedia technology) from the posts to build a special dictionary and load them into the jiebaR program to improve the accuracy of word segmentation. Secondly, the third author of this paper conducted the part-of-speech tagging using the jiebaR package in R language. In corpus linguistics, part-of-speech tagging is the process of marking up a word in a sentence (corpus) as corresponding to a particular part of speech such as nouns, verbs, adverbs, adjectives, pronouns (Kupiec, 1992). All the nouns in Chinese in the text data were retained and sorted to form a set of lexical items.

Phase 3: In-service teachers' participation characteristics in online discourse were analyzed. In addition, the time series analysis was implemented in R language version 3.5.0 using 'xts' packages.

Phase 4: Lexical vector generation. The Chinese lexical items obtained in Phase 2 were sent to Word2vec to produce lexical vectors, which was implemented in R language version 3.5.0 using the approach of Skip-Gram in the text2vec package. After each lexical item was represented as a vector, each post in an online discourse was represented by an average of the vector sum of the lexical items it contains (suppose each post contained n lexical items), see equation 1. The results of this phase were the generation of lexical vectors and post vectors.

$$Vector_{post} = \frac{\sum_1^n Vector_{lexical\ item}}{n} \quad \text{Equation 1}$$

Phase 5: Calculation of new information. Each post was divided into two parts in terms of the information it contained. One part represented the given information and the other part represented the new information. Referring to the method used by Hu et al. (2003), all previous posts of the current post (post A) spanned to a vector space (see equation 2). The given information in post A was represented by the projection of the vector of post A onto the vector space, see equation 3. The result of equation 3 was a numerical value, which represented the given information contained in post A.

$$\text{Vector space} = \text{span}\ \{\overrightarrow{p_1}, \overrightarrow{p_2}, \ldots, \overrightarrow{p_{t-1}}\} \quad \text{Equation 2}$$

$$Information_{Given} = \|Projection_{vector\ space}(\overrightarrow{p_t})\| \quad \text{Equation 3}$$

Equation 3 is the norm of the projected vector. For example, in order to calculate the given information of the Post 3, the first two posts (post 1 and post 2) were spanned to a vector space (see Figure 4), and the projection of the vector of the post 3 onto the vector space represents the given information of post 3.

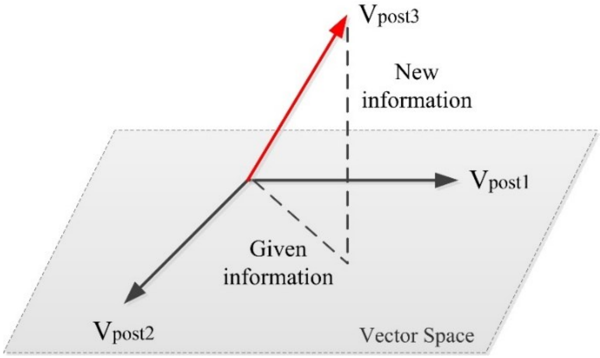


Figure 4. Projecting the vector of the third post to a vector space

The new information was represented by the projection of the vector of post 3 onto the Orthogonal Complement Space of the Vector Space (OCSVS), see equation 4. The result of equation 4 was a numerical value, which represented the new information contained in post 3.

$$Information_{New} = \|Projection_{OCSVS}(\overrightarrow{p_t})\| \quad \text{Equation4}$$

Then, the new information contained in each post was standardized, see equation 5.

$$\text{New}\ (post_i) = \frac{Information_{New}}{Information_{Given} + Information_{New}} \quad \text{Equation 5}$$

The mean value, standard deviation of the new information contained in every post and the evolution of the new information over time were also calculated. The new information analysis was implemented in R language version 3.5.0 using Matrix and limSolve packages. The Matrix package contains numerous methods for and operations on the matrix, including triangular, symmetric, and diagonal matrices, both dense and sparse and with pattern, logical and numeric entries (Bates & Maechler, 2019). The limSolve package was used to solve the linear inverse matrix of a given matrix (Soetaert, van den Meersche, & van Oevelen, 2017).

Phase 6: Calculation of topic relevance. The correlation between a post vector and the document vector (an average of the vector sum of all the posts in the discourse) indicated the extent to which the post was related to the content of the entire online discourse (Dascalu, Trausan-Matu, Dessus, & Mcnamara, 2015). Suppose an online discourse contained m posts, the representation of the document vector is shown in equation 6.

$$Vector_{document} = \frac{\sum_1^m Vector_{post}}{m} \quad \text{Equation 6}$$

The calculation of the correlation between the post vector and the document vector is shown in equation 7.

$$\text{Correlation} = \cos(\theta) = \frac{\boldsymbol{V_{post}} \cdot \boldsymbol{V_{document}}}{\|\boldsymbol{V_{post}}\|\|\boldsymbol{V_{document}}\|} = \frac{\sum_{i=1}^{n} \boldsymbol{V_{post}}_i \boldsymbol{V_{document}}_i}{\sqrt{\sum_{i=1}^{n} \boldsymbol{V_{post}}_i^2}\sqrt{\sum_{i=1}^{n} \boldsymbol{V_{document}}_i^2}} \quad \text{Equation 7}$$

In equation 7, cos represents the cosine distance between two vectors, which was widely used to calculate the correlation degree of two vectors. Next, the mean and standard deviation of the topic relevance of all posts and the evolution of the topic relevance over time were analyzed.

Phase 7: Identification of the cluster of posts and in-service teachers' online discourse characteristics. Before the cluster analysis, the values of new information and topic relevance were standardized, which was a common step in cluster analysis (Hastie, Tibshirani, & Friedman, 2013). Many methods can be used to implement cluster analysis, and the agglomerative hierarchical clustering method was adopted in this study for solving similar problems (e.g., Kovanović, Gašević, Joksimović, Hatala, Adesope, 2015; Wise, Speer, Marbouti, Hsiao, 2013). The cluster analysis procedure was implemented in R language version 3.5.0 using 'stats' packages. The process of cluster analysis included two steps: the first step was to determine the optimal number of clusters, and the second step was to implement cluster analysis. Two indicators, the cluster centroids and the average silhouette index, were evaluated to select the optimal number of clusters. The clustering dendrogram is shown in Figure 5. When implementing cluster analysis, the agglomerative hierarchical clustering method was used with the

Euclidean distance and Ward's agglomeration criteria. After determining the cluster of each post, we further identified in-service teachers' online discourse characteristics.

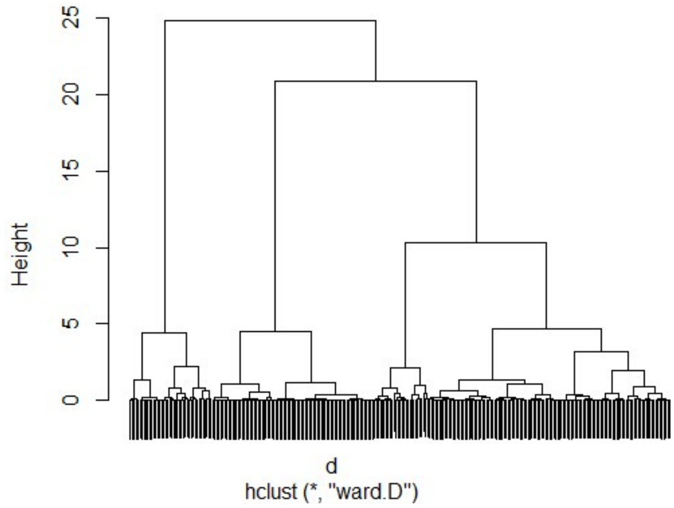


Figure 5. Dendrogram of post clustering

5. Results

5.1. Algorithm evaluation

In order to ensure the validity and reliability of the proposed algorithm, one coder in educational technology and one in teacher professional development independently rated each post on a five-point Likert scale ranging from 1 (lowest) to 5 (highest) on two indicators (new information and topic relevance). Coding results showed that the inter-rater reliability Kappa was 0.667 ($p < .01$), which demonstrated fair to good reliability. Then, two coders discussed the differences in codes and combined the results. Cohen's Kappa coefficient (K) was calculated to measure the agreement between the proposed algorithm and human in two indicators. The results showed the moderate agreement rate between the proposed algorithm and human: the coefficient (K) of new information was *0.618* and the coefficient (K) of topic relevance was *0.775*.

5.2. What were the in-service teachers' participation characteristics in online discourse?

The average contribution of each in-service teacher was 4.59 posts, and the average length of posts was 39.47 characters in Chinese. Figure 6 shows the temporal characteristics of the teachers' online discourse. In-service teachers contributed a large number of posts in the early stages of online discourse. As time went by, the number of posts contributed by the in-service teachers decreased. By the end of the online discourse, in-service teachers' contribution increased slightly.

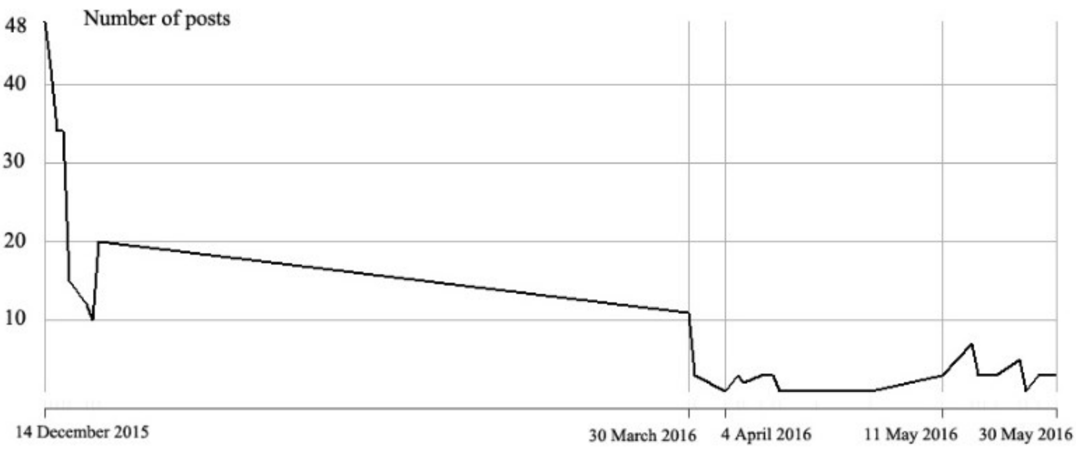


Figure 6. Change of the number of posts over time

5.3. How much new information was contained in each of the in-service teacher's post in online discourse?

In-service teachers of the mathematics community contributed a total of 275 posts, and the average length of posts was 41.138 characters in Chinese. The minimum value of new information (= 0) indicated that some posts did not contain any new information, while the maximum value (= 1) indicated that the information contained in a post was entirely new information. The mean of new information contained in the posts was 0.221.

Figure 7 further shows how the new information contained in the posts changed over time. The dotted line in Figure 7 represented the mean amount of new information. In-service teachers' posts contained much new information in the early stages of online discourse. As time went on, the new information contained in the posts gradually decreased. It was worth noting that posts with new information of 0 appeared at all stages of the online discourse and appeared intensively in the later stage of the online discourse.

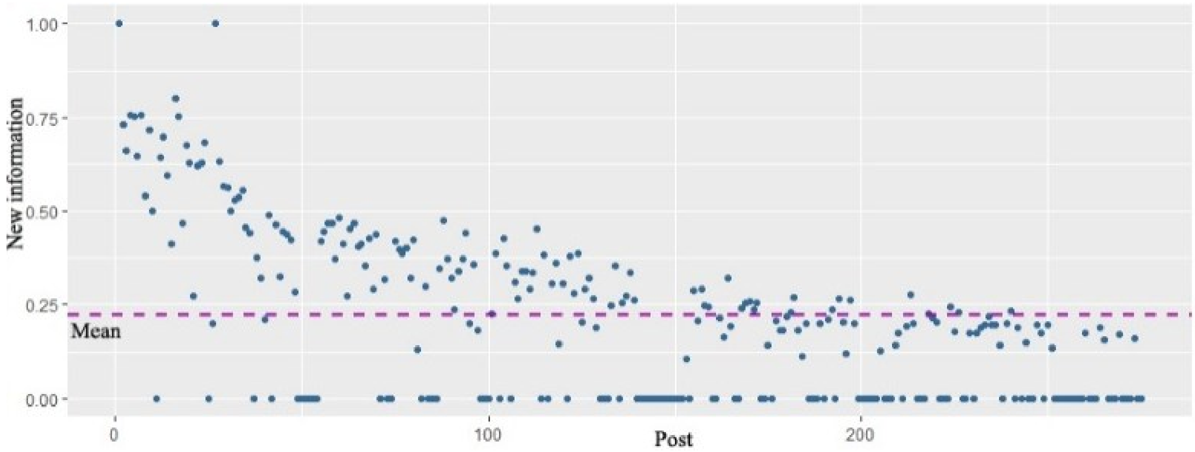


Figure 7. Change of the amount of new information over time

5.4. How relevant were the in-service teacher's posts to the discussion topic in online discussion discourse?

The minimum value of topic relevance (= 0) indicated that a post had nothing to do with the discussion topic, while the maximum value (= 0.943) indicated that the content of a post was highly relevant to the discussion topic. The mean of relevance was 0.708, indicating that most posts were very relevant to the discussion topic.

Figure 8 further shows how the topic relevance changed over time. The dotted line in Figure 8 represents the mean of the topic relevance. The content of in-service teachers' posts was highly related to the discussion topic, which appeared in all stages of online discourse. It was rare that the content of in-service teachers' posts was completely unrelated to the discussion topic.

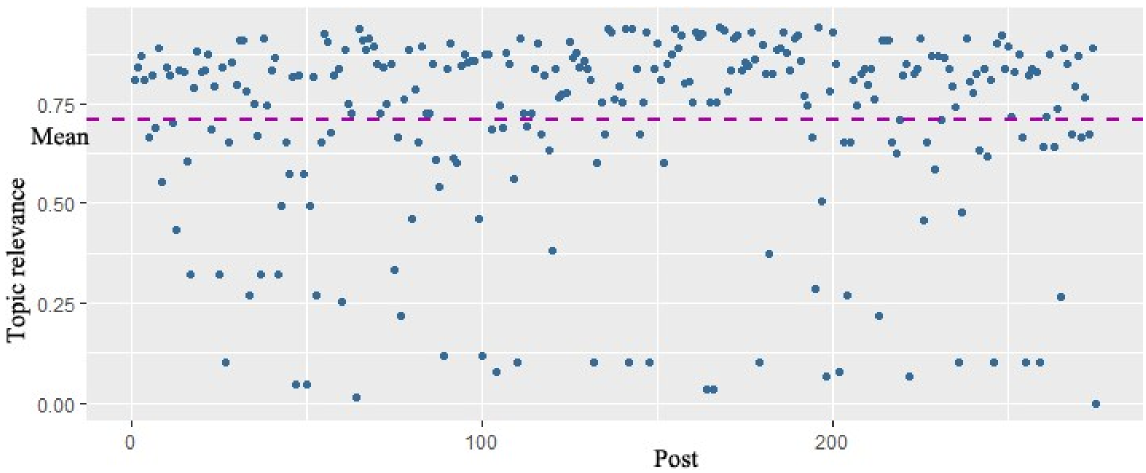


Figure 8. Change of the topic relevance over time

5.5. What were the in-service teachers' online discourse characteristics in terms of cognitive aspect?

In order to determine the optimal number of clusters, we compared the differences of the cluster centroids and the average Silhouette index of different clustering solutions. Although the two-cluster solution had the highest average Silhouette index, we decided to use the three-cluster solution, where one cluster of the two-cluster solution was divided into two smaller sub-clusters for the following reasons. First, in the two-cluster solution, one cluster included 83.6% of the posts (see Table 2) and a basic distinction between related or not related to the discussion topic which had very little practical significance. Second, we can see the difference between the two sub-clusters from the cluster centroids of the three-cluster solution, which are associated with in-service teachers' cognitive characteristics and were aligned with literatures on learners' cognitive engagement (e.g., Ding, Er, Orey, 2018; Ding, Kim, & Orey, 2017). Based on the above considerations, this study used the three-cluster solution (see Table 3).

Table 2. Characteristics of the two-cluster solution

Cluster #	Cluster centroids (Relate, New)	Number of posts (%)	Characteristics
Cluster 1	(0.803, 0.220)	230 (83.6%)	High relevance, little new information
Cluster 2	(0.221, 0.221)	45 (16.4%)	Low relevance, little new information

Table 3. Characteristics of the three-cluster solution

Cluster #	Cluster centroids (Relate, New)	Number of posts (%)	Characteristics
Cluster 1	(0.764, 0.484)	75 (27.3%)	High relevance, much new information
Cluster 2	(0.810, 0.104)	160 (58.2%)	High relevance, little new information
Cluster 3	(0.193, 0.190)	40 (14.5%)	Low relevance, little new information

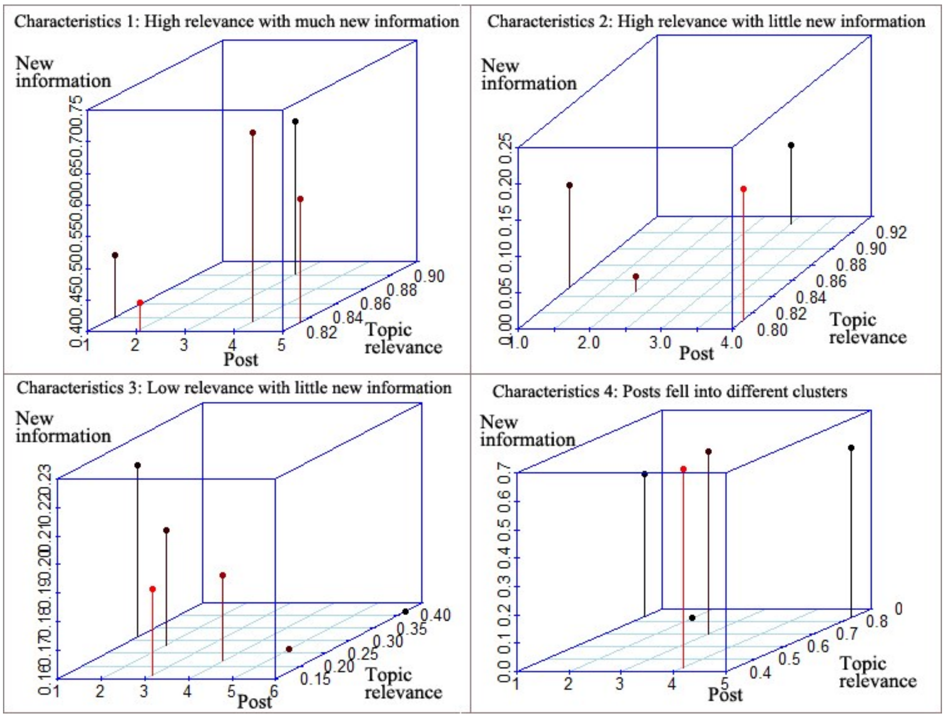


Figure 9. Four types of online discourse characteristics

Combining the results of Figures 7 and 8, it is easy to observe the following results: cluster 1 appeared primarily in the early stage of online discourse; cluster 2 existed in the whole process of online discourse; and cluster 3 occasionally appeared in the online discourse. Four typical types of in-service teachers' online discourse characteristics are shown in Figure 9. The in-service teachers with the first type of characteristics always contributed posts that were related to online discussion topics and contained much new information. However, the in-service teachers with the second type of characteristics always contributed posts that were related to online discussion topics but with little new information. In-service teachers with the third type of characteristics contributed many posts, but most of the posts were of low topic relevance and little new information. In terms of the posts contributed by in-service teachers with the fourth type of characteristics, some posts were highly related to the discussion topic but with zero new information, while some with low topic relevance but much new information included.

6. Discussion

In-service teachers contributed an average of 4.59 posts and the average length of each post was 39.47 characters in Chinese. Moreover, in-service teachers' contribution mainly appeared in the early stage of online discourse. This finding was inconsistent with other studies. In Liu's (2012) study, an online video-case discussion community was used to foster in-service teachers' professional development and the results showed that in-service teachers posted an average of 13.72 messages per person. In another study, teachers' contribution across the 6 weeks of online synchronous discussion was quite consistent (Chen, Chen, & Tsai, 2009). Two possible reasons may be related to in-service teachers' low and unbalanced contribution. The first was the scoring rules of the teacher professional development program. Many in-service teachers contributed 5 posts in succession in order to complete the learning tasks as soon as possible, and stopped posting once the basic learning requirements were met. The second reason was due to in-service teachers' busy work schedule. When faced with the contradiction between work and learning, in-service teachers often expected to finish their learning task as early as possible. Due to the constant changes of participants, a good discourse culture in the online professional learning community had not been formed and it was difficult for in-service teachers to have high-level reflection and collaborative knowledge construction (Lee & Brett, 2015).

New information contained in learners' contribution is conductive to the accumulation and continuation of ideas (Keles, 2018), as well as the promotion of collaborative knowledge building. The results indicated that the online discourse activity presented in this study had encouraged and facilitated in-service teachers' active and constructive contributions related to classroom practice to some extent. In-service teachers' posts contained much new information in the early stages of the online discourse, while in the later stage of the online discourse, the amount of new information became very small. From a technical perspective, as the online discourse progressed, most of the information related to the topic had been discussed. If an in-service teacher repeated something similar, there would be little new information in the post. However, from a pedagogical perspective, this situation should be avoided. An effective online professional learning community relies on the continuous, active participation and contribution of team members to solve emerging problems and generate new knowledge. From this point of view, a good community atmosphere had not been formed in the mathematics community. Two possible reasons may be related to this situation. On the one hand, although online discourse environments facilitated collaboration among in-service teachers, they still lacked necessary discussion skills, such as reviewing existing information, or comparing inconsistencies between posts. On the other hand, the online discourse platform could not give hints to in-service teachers about what had been discussed. Therefore, the necessary scaffolding and technical support can help the formation and development of the online professional learning community (e.g., Marbouti & Wise, 2016).

Whether the content of posts was related to the discussion topic not only influences the cohesion of the online discourse, but also influences the shared culture of the online learning community (Tsiotakis & Jimoyiannis, 2016). The results showed that most of in-service teachers' posts were relevant to the discussion topic, indicating that the cohesion of in-service teachers' online discourse was high. This finding was correlated with previous studies. The teachers in the online professional learning community actively exchanged information, and provided and sought advices (e.g., Lantz-Andersson et al., 2018; Trust, Krutka, & Carpenter, 2016). In an organized online discourse, teachers concentrated on the discussion topic much more frequently (e.g., Keles, 2018). This behavior is of value to the online professional learning community, as it indicated that in-service teachers were actively thinking about the topic of online discourse and related it with their own professional knowledge and experience.

The results of cluster analysis showed three different groups of posts. The second group of posts, those highly related to the discussion topic but containing little new information, accounted for the largest number. This group of posts appeared mainly in the middle and latter stages of the online discourse. The first group of posts, which closely related to the discussion topic and containing much new information, came in the second place. This group of posts appeared mainly in the early stage of the online discourse. The posts that were less relevant to the discussion topic and contained little new information appeared sporadically in the online discourse. The results of the cluster analysis suggested that in-service teachers' behaviors and the extent that they contributed new knowledge might be influenced by teachers' tight schedules, the activity design of the online discussion, and the support strategies of the online learning platform. Therefore, it is necessary to design different intervention strategies for in-service teachers with different cognitive characteristics in the OPLCs. For example, for teachers who consistently contribute posts that are highly relevant to the topic and contain a lot of new information, more rewards and opportunities to show their prestige are necessary. For in-service teachers who consistently contribute posts that are unrelated to the topic and have no new information, they need to be prompted. In addition, for teachers who consistently contribute posts that are relevant to the topic but contain less new information, they need to be encouraged to read their colleagues' posts and contribute innovative ones, for instance.

This study had two major contributions. The first one was that it proposed a completely automatic evaluation approach for detecting in-service teachers' cognitive engagement in the OPLC. The method used in this study was highly automatic. It helped introduce an intelligent supervision system in the online learning community to guide and intervene learners' online discourse. If learners contribute little new information (e.g., always repeating what has been discussed), community information sharing culture, problem solving process, and collaborative knowledge construction will be affected. At this point, it is necessary for the online learning platform to remind learners of what is being discussed and encourage learners to think from different perspectives. On the other hand, if the posts contributed by learners are off-topic or the topic relevance is very low, it will affect the cohesion of the discussion among community members. At this point, it is necessary for the online learning platform to intelligently determine the topic relevance of the posts. The method presented in this paper established a technical basis for the implementation of the intervention. The second was that it identified in-service teachers' online discourse characteristics from the cognitive aspects of interaction, which was helpful to identify the learner roles emerging in online learning settings, such as core members (those who actively contribute new information and have a high degree of topic relevance) and marginal roles (those who contribute a small amount of posts and have little new information). Further design of intervention strategies for different roles, such as incentive strategies for core members and prompt strategies for marginal members, will help core and marginal members of the online learning communities to better monitor and regulate the online learning process.

This study had a few important implications. First, when evaluating in-service teachers' learning engagement in an online learning community, we should pay attention to both the quantity and quality of their contribution. Previous studies used teachers' behaviors, such as the number of posts, replies, and resources downloaded to represent teachers' learning engagement in online discussions and did not notice that many posts with no new information appeared (Xing & Gao, 2018). Therefore, just counting in-service teachers' behaviors did not represent their real performance in online discussions. Second, this study had important implication for building an OPLC. Since in-service teachers in the OPLC had different online discourse characteristics, teacher educators need to design different supervision and intervention strategies. Third, this study provided a bridge between a theoretical description of leaner's interaction in the online learning environment and a computational method of the interaction. Alternative analyses or interpretations were definitely possible given the types of analysis conducted in this study. For example, the computational method could be used to evaluate the performance of different online discussion pedagogy in online learning settings (e.g., MOOCs), or the learning performance of different learning groups under the same online discussion pedagogy.

7. Conclusion, limitations, and future study

This study integrated learning analytics and educational data mining to build a workflow for solving a key issue in the OPLC: to detect in-service teachers' cognitive engagement in online discussion activity. Results showed that in-service teachers' posts contained more new information in the early stages of online discourse and the average amount of new information contained in each post was 0.221. When using the calculation method of topic relevance proposed in this paper, most of in-service teachers' posts were highly relevant to the discussion topic. In addition, cluster analysis generated three different groups of posts with unique characteristic. Based on the three groups of posts, it was easy to identify in-service teachers' online discourse characteristics. The workflow and research results presented in this study were not only applicable to the discourse analysis in the OPLC, but also contribute to the establishment of an intelligent online learning environment (e.g., MOOCs) which can automatically monitor and provide feedback on learners' discourse.

There were two major limitations in this study. Firstly, the research method proposed in this paper tended to ignore some small patterns that may be important. For example, the posts that were not highly relevant to the discussion topic but contained a lot of new information were ignored. In fact, such posts may contain innovative ideas that were simply not appreciated by the community members. Secondly, this study only analyzed the online discourse characteristics of a mathematics community. When researchers and teachers apply the conclusions of this study, they need to consider the context of the study. Future research directions are to: (1)

explore the relationship between the quantity of new information in online discourse and in-service teachers' perception, collaborative learning quality and learning score; (2) design different intervention strategies for in-service teachers with different types of online discourse characteristics; and (3) deeply understand the factors influencing in-service teachers' online discourse characteristics through interviews and questionnaire surveys.

Acknowledgement

This paper was supported by the National Natural Science Foundation of China (Grant No. 62077016). This work was also conducted as part of a research project that was funded by the School of Educational Information Technology, Central China Normal University of China (Project No. CCNUSEIT202008).

Statements on open data, ethics and conflict of interest

Data can be accessed by contacting the author (saved in a personal repository). Ethical approvals were gained from the hosting institution. This research has no conflicts of interest.

References

Atapattu, T., Thilakaratne, M., Vivian, R., & Falkner, K. (2019). Detecting cognitive engagement using word embeddings within an online teacher professional development community. *Computers & Education*, *140*, 103594.

Baran, B., & Cagiltay, K. (2010). The Dynamics of online communities in the activity theory framework. *Educational Technology & Society*, *13*(4), 155-166.

Bates, D., & Maechler, M. (2019). *Matrix: Sparse and Dense Matrix Classes and Methods*. Retrieved from https://cran.r-project.org/web/packages/Matrix/index.html

Bing, L., Niu, Z.-Y., Li, P., Lam, W., & Wang, H. (2018). Learning a unified embedding space of web search from large-scale query log. *Knowledge-Based System*, *150*, 38-48.

Cerisara, C., Král, P., & Lenc, L. (2018). On the effects of using Word2vec representations in neural networks for dialogue act recognition. *Computer Speech & Language*, *47*, 175-193.

Chafe, W. L. (1976). Givenness, contrastiveness, definitensess, subjects, topics, and point of view. In C. N. Li (Eds.), *Subject and Topic* (pp. 25-55), New York, NY: Academic Press.

Chen, J.-M., Chen, M.-C., & Sun, Y., S. (2010). A Novel approach for enhancing student reading comprehension and assisting teacher assessment of literacy. *Computers & Education*, *55*(3), 1367-1382.

Chen, Y., Chen, N.-S., & Tsai, C.-C. (2009). The use of online synchronous discussion for web-based professional development for teachers. *Computers & Education*, *53*(4), 1155-1166.

Chen, H.-L., Fan, H.-L., & Tsai, C.-C. (2014). The Role of community trust and altruism in knowledge sharing: An Investigation of a virtual community of teacher professionals. *Educational Technology & Society*, *17*(3), 168-179.

Cho, H. (2016). Under co-construction: An Online community of practice for bilingual pre-service teachers. *Computers & Education*, *92-93*, 76-89.

Dascalu, M., Trausan-Matu, S., Dessus, P., & McNamara, D. (2015). Discourse cohesion: A Signature of collaboration. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 350-354). New York, NY, United States: Association for Computing Machinery. doi:10.1145/2723576.2723578

Denzin, N. K., & Lincoln, Y. S. (1994). *Handbook of qualitative research*. Thousand Oaks, CA: Sage Publications.

Ding, L., Er, E., & Orey, M. (2018). An Exploratory study of student engagement in gamified online discussions. *Computers & Education*, *120*, 213-226.

Ding, L., Kim, C., & Orey, M. (2017). Studies of student engagement in gamified online discussions. *Computers & Education*, *115*, 126-142.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59-109.

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, *2*(2-3), 87-105.

GaŠević, D., Joksimović, S., Eagan, B. R., & Shaffer, D. W. (2019). SENS: Network analysis to combine social and cognitive perspectives of collaborative learning. *Computers in Human Behavior*, *92*, 562-577.

Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research*, *17*(4), 397-431.

Halliday, M. A. K. (1967). Notes on transitivity and theme in English. *Journal of Linguistics, 3*, 199-244.

Harris, Z. (1952). Discourse analysis. *Language*, *28*(1), 1-30.

Hashimoto, T. B., Alvarez-Melis, D., & Jaakkola, T. S. (2015). Word, graph and manifold embedding from Markov processes. *Computing Research Repository.* Retrieved from https://arxiv.org/pdf/1509.05808.pdf

Hastie, T. J., Tibshirani, R. J., & Friedman, J. H. (2013). *The Elements of statistical learning: Data mining, inference, and prediction*. New York, NY: Springer.

Heilporn, G., & Lakhal, S. (2020). Investigating the reliability and validity of the community of inquiry framework: An Analysis of categories within each presence. *Computers & Education*, *145*, 103712.

Hu, X., Cai, Z., Louwerse, M., Olney, A., Penumatsa, P., & Graesser, A. (2003). A Revised algorithm for latent semantic analysis. In *Proceedings of the 18th international joint conference on Artificial intelligence* (pp. 1489-1491). San Francisco, CA: Morgan Kaufmann Publishers Inc.

Keles, E. (2018). Use of Facebook for the community services practices course: Community of inquiry as a theoretical framework. *Computers & Education*, *116*, 203-224.

Kim, D., Park, Y., Yoon, M., & Jo, I.-H. (2016). Toward evidence-based learning analytics: Using proxy variables to improve asynchronous online discussion environments. *Internet and Higher Education*, *30*, 30-43.

Kovanović, V., Gašević, D., Joksimović, S., Hatala, M., & Adesope, O. (2015). Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions. *The Internet and Higher Education*, *27*, 74-89.

Kupiec, J. (1992). Robust part-of-speech tagging using a hidden MarKov model. *Computer Speech & Language*, *6*(3), 225-242.

Lantz-Andersson, A., Lundin, M., & Selwyn, N. (2018). Twenty years of online teacher communities: A Systematic review of formally-organized and informally–developed professional learning groups. *Teaching and Teacher Education*, *75*, 302-315.

Lee, K., & Brett, C. (2015). Dialogic understanding of teachers' online transformative learning: A Qualitative case study of teacher discussions in a graduate-level online course. *Teaching and Teacher Education*, *46*, 72-83.

Li, J., Li, J., Fu, X., Masud, M. A., & Huang, J. Z. (2016). Learning distributed word representation with multi-contextual mixed embedding. *Knowledge-Based Systems*, *106*, 220-230.

Liu, M.-H. (2012). Discussing teaching videocases online: Perspectives of preservice and inservice EFL teachers in Taiwan. *Computers & Education*, *59*(1), 120-133.

Marbouti, F., & Wise, A. F. (2016). Starburst: A New graphical interface to support purposeful attention to others' posts in online discussions. *Educational Technology Research and Development*, *64*, 87-113.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems*. Red Hook, NY, United States: Curran Associates Inc. Retrieved from https://arxiv.org/pdf/1310.4546.pdf

Peng, X., & Xu, Q. (2020). Investigating learners' behaviors and discourse content in MOOC course reviews. *Computers & Education*, *143*, 103673. doi:10.1016/j.compedu.2019.103673

Polizzi, S., Head, M., Barrett-Williams, D., Ellis, J., Roehrig, G. H., & Rushton, G. T. (2018). The Use of teacher leader roles in an online induction support system. *Teaching and Teacher Education*, *75*, 174-186.

Prestridge, S. (2019). Categorising teachers' use of social media for their professional learning: A Self-generating professional learning paradigm. *Computers & Education*, *129*, 143-158.

Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical Pragmatics* (pp. 223-255), New York, NY: Academic Press.

Sinoara, R. A., Camacho-Collados, J., Rossi, R. G., Navigli, R., & Rezende, S. O. (2019). Knowledge-enhanced document embeddings for text classification. *Knowledge-Based Systems*, *163*, 955-971.

Soetaert, K., van den Meersche, K., & van Oevelen, D. (2017). *limSolve: Solving Linear Inverse Models*. Retrieved from https://cran.r-project.org/web/packages/limSolve/

Sung, Y.-T., Liao, C.-N., Chang, T.-H., Chen, C.-L., & Chang, K.-E. (2016). The Effect of online summary assessment and feedback system on the summary writing on 6th graders: The LSA-based technique. *Computers & Education*, *95*, 1-18.

Trust, T., Krutka, D. G., & Carpenter, J. P. (2016). “Together we are better”: Professional learning network for teachers. *Computers & Education*, *102*, 15-34.

Tsai, I.-C. (2012). Understanding social nature of an online community of practice for learning to teach. *Educational Technology & Society*, *15*(2), 271-285.

Tsiotakis, P., & Jimoyiannis, A. (2016). Critical factors towards analyzing teachers’ presence in online learning communities. *The Internet and Higher Education*, *28*, 45-58.

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

Wang, X., Wen, M., & Rosé, C. P. (2016). Towards triggering higher-order thinking behaviors in MOOCs. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (pp. 398–407). New York, NY, United States: Association for Computing Machinery. doi:10.1145/2883851.2883964

Wise, A. F., Speer, J., Marbouti, F., & Hsiao, Y. -T. (2013). Broadening the notion of participation in online discussions: Examining patterns in learners’ online listening behaviors. *Instructional Science*, *41*(2), 323-343.

Wu, P., Yu, S., & Wang, D. (2018). Using a learner-topic model for mining learner interests in open learning environments. *Educational Technology & Society*, *21*(2), 192-204.

Xie, K., Heddy, B. C., & Greene, B. A. (2019). Affordances of using mobile technology to support experience-sampling method in examining college students’ engagement. *Computers & Education*, *128*, 183-198.

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

Yang, S.-H. (2016). Conceptualizing effective feedback practice through an online community of inquiry. *Computers & Education*, *94*, 162-177.

Zhang, S., Liu, Q., Chen, W., Wang, Q., & Huang, Z. (2017). Interactive networks and social knowledge construction behavioral patterns in primary school teachers’ online collaborative learning activities. *Computers & Education*, *104*, 1-17.

Zhang, D., Xu, H., Su, Z., & Xu, Y. (2015). Chinese comments sentiment classification based on Word2vec and SVM. *Expert Systems with Applications*, *42*(4), 1857-1863.

Ioannou, A., Bhagat, K. K., & Johnson-Glenberg, M. C. (2021). Guest Editorial: Learning Experience Design: Embodiment, Gesture, and Interactivity in XR. *Educational Technology & Society*, *24* (2), 73–76.

Guest Editorial: Learning Experience Design: Embodiment, Gesture, and Interactivity in XR

Andri Ioannou[1*], Kaushal Kumar Bhagat[2] and Mina C. Johnson-Glenberg[3]

[1]Cyprus Interaction Lab, Department of Multimedia and Graphic Arts, Cyprus University of Technology and Research Centre on Interactive Media, Smart Systems and Emerging Technologies, Cyprus // [2]Centre for Educational Technology, Indian Institute of Technology Kharagpur, India // [3]Arizona State University, Department of Psychology and Embodied Games, Tempe, AZ, USA // andri.i.ioannou@cut.ac.cy // kkntnu@hotmail.com // minaj@embodied-games.com

[*]Corresponding author

ABSTRACT: The concepts of embodiment and embodied learning are gaining traction in the field of education. This special issue aims to synthesize current knowledge on the design and evaluation of learning in immersive and embodied learning environments, mediated by XR (eXtended Reality) technologies. Of the 14 invited submissions, six (6) were finally accepted for publication. The collection of works in this special issue provides insights on best practices for learning experience design, based on systematic or empirical data and analysis on learning outcomes or processes.

Keywords: XR, AR, VR, MR, Extended reality, Virtual reality, Augmented reality, Mixed reality, Embodied learning, Immersive learning, Learning environments

1. Introduction

The concepts of embodiment and embodied learning, deeply rooted in theories of embodied cognition, are gaining traction in the field of education. New educational technologies enable researchers and practitioners to include more gestures and body movements into their learning experience design, creating immersive and gesture-rich learning environments. Such embodied environments should enable multi-modal and multi-sensory forms of interaction through gestures and bodily movement, tactile, and auditory sensory experiences. While the interplay of new forms of technology and learning is complex, recent evidence suggests that learning experience design, pedagogy, and practice with embodied and immersive learning technologies can have important effects on learning, engagement, and achievement in multiple educational settings, including formal and non-formal (Georgiou & Ioannou, 2019). This special issue aims to synthesize current knowledge on the design and evaluation of learning in immersive and embodied learning environments.

The specific scope of this special issue is to publish research that addresses learning in immersive and embodied environments mediated by XR (eXtended Reality) technologies, an umbrella term that refers to the spectrum of AR, VR, and mixed-reality environments, immersive gaming environments, immersive escape rooms, XR mobile computing platforms, etc. The focus of this special issue is not the technology per se but rather issues related to learning experience design (or learning design), the process continuum of teaching, learning and assessment, and how these are affected or enhanced by XR technologies. We have sought out evidence-based educational applications and research that meshes pedagogy and practice in all types of learning environments. Indeed, empirical research on the intersection of pedagogy and practice of XR learning in authentic settings is very limited to date. The aim of this special issue is to provide insights on best practices for learning experience design, based on systematic or empirical data and analysis on learning outcomes or processes.

The initial call for paper proposals resulted in 40 high caliber mini prospectuses, of which 14 were invited to submit full papers. All the invited submissions underwent ET&S's rigorous blind peer review process, with over 30 field specialists from around the world agreeing to provide expert reviews. Of the 14 invited submissions, six (6) were finally accepted for publication. The editorial team allowed the more robust and well-developed submissions to be processed for publications after two rounds of revisions and extensive work by the authors and reviewers to ensure high quality.

2. Overview of papers

The variety of technologies, research methods, and research contexts in our collection of papers allows the reader of this special issue to gain a comprehensive insight into the current state-of-the-art in learning experience design in XR.

Kang, Diederich, Lindgren and Junokas (2021) focus on gesture patterns and learning in an embodied XR science simulation. The innovative ELASTIC3S system allows learners to interact with different science simulations via whole-body gestures (e.g., hand waving, kicking). The authors found trends in the use of the science simulations directly linked to students' struggles in understanding the underlying ideas or use of the system, as well as with their learning performance. Their findings can inform the design of the real time assistance within the embodied simulation, thus enabling an adaptive learning experience.

In their work, Birt and Vasilevski (2021) examine immersive VR learning in the context of building information modeling (BIM) in architecture, engineering, and construction. The researchers experimented with a multiuser (synchronous and collaborative) vs. a single-user (asynchronous) learning environment and found that the learners' experience was better in the multiuser condition. Considering these findings, the authors discuss the future design of immersive virtual reality environments for learning.

Holly, Pirker, Resch, Brettschuh and Gütl (2021) elaborate on the challenging job of designing educational VR platforms to meet the expectations of educators and students. The reflections are the result of the authors' work on the Maroon platform -- a VR environment for teaching physics -- which has been in development for over five years. The authors present recommendations related to immersion, costs, time restrictions, and the learning process to overcome current challenges for learning and teaching with VR in the physics domain.

In their work, Lyons and Mallavarapu (2021) take a new approach to understand VR immersive experiences. They define the concept of collective usability, as the degree to which a group of simultaneous users can make use of an interactive experience where the human-computer and human-human interactions combine to form a complex system. The authors present a simple agent-based model simulation to explore how changes in the number of simultaneous users and the duration, size, and number of the proffered interactives can affect the collective usability of XR learning environments.

Chaker, Binay, Gallot and Hoyek (2021) examine the user experience of the learners in a 3D interactive human anatomy tool. The paper presents the author's research and development approach, which includes two phases of data collection and UX evaluation. The authors found a correlation between student's UX and both their anatomy scores and motor imagery abilities (i.e., their ability to imagine a human movement without any real movement).

Yiannoutsou, Johnson and Price (2021) extend the discussion of the pedagogical design of embodied mathematical experiences for visually impaired children. They present an iterative, design-based case study with visually impaired children to inform the pedagogical design of embodied mathematical experiences. Their system provides opportunities for grounding mathematical ideas in audition and bodily experience. In their work, sensorimotor interaction (e.g., touching a small tangible grid to understand how to then locomote in an auditorily enhanced CAVE with grid-like components) seems particularly promising for visually impaired children. Their findings show how bodily movement and positioning can effectively foster visually impaired children's engagement, and they discuss prerequisites for the implementation of immersive VR in the classroom.

3. Summary and future directions for XR learning experience design research

XR in education has great potential for research and development. XR technologies can be integrated in the learning environment to allow learners to interact with critical elements in a domain without real risk; it can make the "unseen be seen" in ways that 2D media cannot. Using XR technologies, one may enable simulated environmental and socio-cultural interactions between students, educators, practitioners, patients, workers, or other stakeholders in a safe learning environment that authentically simulates the situation, the risks, and the opportunities for action. This special issue aimed to publish state-of-the-art research that uses the principles of embodied cognition meshed with the affordances of XR to engage learners and promote learning outcomes in formal and informal learning environments.

This special issue has compiled research in this burgeoning area; however, it only contributes a small sample of research in the area of learning experience design in XR learning environments. Namely, while this special issue addresses different learners (i.e., high school students, special education) and types of educational context (STEM classroom, design, physiology), there is still a need for more studies. For example, these studies might include younger students, use a larger variety of formal and informal learning situations, different research domains (e.g., humanities and empathy induction), and perhaps explore the boundaries of interactivity in three dimensions (e.g., where is the tipping point for too much interactivity such that it becomes overwhelming for the learner? What are best scaffolding methodologies?). Moreover, in the scope of learning experience design in XR learning environments, the technologies presented in the various contributions of this special issue include VR environments and 3D simulations of low to medium immersion according to the Johnson-Glenberg taxonomy of embodiment in education (2018). We note that XR also includes mobile AR/VR and highly immersive virtual, augmented, and mixed reality environments. The technology is available (e.g., Hololens) to exploit for the design of highly immersive learning environments using overlays, and several companies are now releasing easy-to-use or “no-code” editors for AR and in-headset VR experiences. Educators are encouraged to experiment with these editors and technologies that allow students to be creators in their classrooms.

In terms of methods, there are both quantitative and qualitative studies in this special issue. The editors would suggest that Design-Based Research (DBR) is a promising methodology to support researchers in the advancement of XR learning design in these early stages of work and experimentation (e.g., Ioannou & Ioannou, 2020). DBR shifts the focus to the design problem or learning need and should help researchers examine the appropriate use of XR technologies to address learning needs. DBR can lead to theoretical understanding and transferable design principles for promoting learning in XR learning environments beyond the context of a particular learning need. Such theoretical understandings or transferable design principles illustrating how XR can enrich the learning environment would help researchers and practitioners communicate with each other and advance XR learning design.

This special issue is a concrete step in uniting the divergent XR and learning communities; it also raises many questions. Some of the studies included in this special issue are still in the early stages and there is undoubtedly much space for extending the work of these authors. Evidently, there is a need for more research that will continue to contribute to the growing empirical literature on learning experience design, pedagogy, and practice with embodied and immersive XR technologies. A framework for XR learning design focused on presenting design principles for learning in XR environments becomes of paramount importance as XR technologies continue to make their way into formal and informal educational settings. Moreover, well-researched developments in this area should begin to find their way to commercialization so that more researchers and practitioners can benefit from using them. As evident in the manuscripts of this special issue, the design, development, and evaluation of XR learning environments is a very demanding job and studies are rarely “perfect.” Leaving these developments on the shelves of research laboratories does not help with the wide dissemination and adoption of important findings in teaching and learning. The opportunities and expected outcomes for the learners can be immediately realized when such developments become commercially available, especially at low cost or for free, and easily accessible to those who might be looking to adopt innovative solutions for teaching and learning. We hold that, XR learning experience design is a fruitful area for research and development going forward. The editors of this special issue hope that this introduction and the six contributions that follow will provide a starting point for further inquiry in this area.

Acknowledgement

This work is partially funded by the European Union’s Horizon 2020 Research and Innovation Programme under Grant Agreement No 739578 and the Government of the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development.

References

Birt, J., & Vasilevski, N. (2021). Comparison of single and multiuser immersive mobile virtual reality usability in construction education. *Educational Technology & Society, 24*(2), 93-106.

Chaker, R., Binay, M., Gallot, M., & Hoyek, N. (2021). User experience of a 3D interactive human anatomy learning tool. *Educational Technology & Society, 24*(2), 136-160.

Georgiou, Y., & Ioannou, A. (2019). Embodied learning in a digital world: a Systematic review of empirical research in K-12 education. In P. Díaz, A. Ioannou, K. K. Bhagat, & M. Spector (Eds.), *Learning in a digital world: A multidisciplinary perspective on interactive technologies for formal and informal education. Springer series: Smart Computing and Intelligence* (pp. 155-177). Singapore: Springer.

Holly, M., Pirker, P., Resch, S., Brettschuh, S., & Gütl, C. (2021). Designing VR experiences – Expectations for teaching and Learning in VR. *Educational Technology & Society, 24*(2), 107-119.

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

Johnson-Glenberg, M. C. (2018). Immersive VR and education: Embodied design principles that include gesture and hand controls. *Frontiers in Robotics and AI, 5*, 81. doi:10.3389/frobt.2018.00081

Kang, J., Diederich, M., Lindgren, R., & Junokas, M. (2021). Gesture patterns and learning in an embodied XR science simulation. *Educational Technology & Society, 24*(2), 77-92.

Lyons, L., & Mallavarapu, A. (2021). Collective usability: Using simulation tools to explore embodied design challenges in immersive, shared mixed-reality experiences. *Educational Technology & Society, 24*(2), 120-135.

Yiannoutsou, N., Johnson, R., & Price, P. (2021). Non-visual virtual reality: Considerations for the pedagogical design of embodied mathematical experiences for visually impaired children. *Educational Technology & Society, 24*(2), 151-163.

Kang, J., Diederich, M., Lindgren, R., & Junokas, M. (2021). Gesture Patterns and Learning in an Embodied XR Science Simulation. *Educational Technology & Society*, *24* (2), 77–92.

Gesture Patterns and Learning in an Embodied XR Science Simulation

Jina Kang[1*], Morgan Diederich[1], Robb Lindgren[2] and Michael Junokas[3]
[1]Instructional Technology and Learning Sciences, Utah State University, USA // [2]Curriculum & Instruction, University of Illinois at Urbana-Champaign, USA // [3]Media and Cinema Studies, University of Illinois at Urbana-Champaign, USA // jina.kang@usu.edu // morgan.wood23@aggiemail.usu.edu // robblind@illinois.edu // junokasm@gmail.com
[*]Corresponding author

ABSTRACT: Recent research has emphasized the importance of leveraging embodied interactions for learning critical STEM concepts. ELASTIC^{3}S—an embodied environment for learning about cross-cutting concepts (i.e., non-linear growth)—allows learners to interact with different science simulations through whole-body gestures. Technological advances in gesture recognition can track and respond to students' gestures, however, there has been little investigation into how the gestures performed in these environments relate to subsequent learning. The need for sequential pattern recognition methods is critical in embodied learning if we are to understand how gestural interaction with a simulation facilitates learning. Using data collected via Microsoft Kinect V2 from twelve college students, we applied multivariate Dynamic Time Warping for clustering to identify gestural patterns in ELASTIC^{3}S as evidence for embodied learning processes. Our findings showed that identified trends of simulation use were indicative of students' struggles to understand the underlying ideas or use of the system and were associated with learning performance. These indicators can potentially be used to leverage real time, in-simulation assistance and promote a more adaptive learning experience via embodied simulations.

Keywords: Embodied learning, XR Science education simulations, Gesture recognition, DTW clustering, Time series analysis

1. Introduction

There has been significant interest in leveraging learners' embodied interactions for teaching critical STEM concepts (Lindgren et al., 2016; Nathan & Walkington, 2017; Stieff et al., 2016). This interest builds upon theories of embodied cognition that assert a fundamental connection between the actions/perceptual processes of the body and how people think and learn (Glenberg, 2010; Shapiro, 2019; Wilson, 2002). Research has shown that learners can be prompted to perform gestures and enact their emerging understanding of STEM ideas in ways that promotes new learning (Gallaher & Lindgren, 2015; Lindgren, 2014). This seems to be true even for abstract ideas and unseen processes such as molecular interactions (e.g., Mathayas et al., 2019). A particularly challenging concept that cuts across STEM domains is non-linear growth: understanding where it is present and how it differs from linear growth (Tretter et al., 2006).

We designed an XR embodied learning environment to target students' ideas about non-linear growth called ELASTIC^{3}S (Embodied Learning Augmented through Simulation Theaters for Interacting with Cross-Cutting Concepts in Science). The ELASTIC^{3}S platform allows learners to control different science simulations with user-defined whole-body gestures (e.g., hand waving, kicking). This particular implementation of XR uses gesture recognition technology paired with multiple large digital displays to create an interactive and immersive environment where students engage with science simulations through gesture. Although these types of environments have shown promise for summative learning outcomes (Johnson-Glenberg et al., 2014; Lindgren et al., 2016), there has been little investigation into how the progression of gestures that learners perform in these environments relate to their subsequent learning, and how the gestures themselves can be used for the purposes of assessment and monitoring. The purpose of this work is to apply pattern recognition algorithms to the gestures that students perform in an embodied XR science education simulation as a means of understanding what is learned and when personalized feedback should be presented.

2. Relevant work

2.1. Embodied learning

Processes of human cognition are deeply rooted in how the body interacts with the environment (Gallagher, 2006). Our understanding of how the world works is organized around the human sensorimotor system and our various modes of perception and action (Barsalou, 2008). Embodiment has increasingly become a focus in various research domains including cognitive psychology (Glenberg, 2010), linguistics (Lakoff & Johnson, 1980), and the performing arts (Noice & Noice, 2006). In education, researchers are applying ideas from embodied cognition to the design of learning environments. "Embodied learning" is essentially the forging of meaningful connections between body movements, artifacts, and learning content (Duijzer et al., 2019; Lindgren & Johnson-Glenberg, 2013; Skulmowski & Rey, 2018). This type of learning is based around the idea that a student has agency and an active role in their learning experience, and that learning activities can be designed to effectively leverage alignments between physical modes of interaction and target concepts. Relevant work has found that the use of embodied learning leads to improved learning outcomes and understanding in multiple domains (e.g., Han & Black, 2011; Glenberg, 2008; Goldin-Meadow, 2011; Segal et al., 2014). Specifically, research has been investigated in STEM education that identified a relationship between gestures and improved scientific reasoning (Crowder, 1996).

Studies have demonstrated the learning effectiveness of embodied learning environments compared to more traditional learning settings. Johnson-Glenberg et al. (2011) conducted an experimental design studying if embodied learning using XR was more effective than traditional classroom learning in which seventy-one 9th graders participated. Results indicated that the embodied environment led to greater knowledge gains. As a follow-up study, they examined if the XR embodied learning environment was more effective than a desktop simulation, and they found that embodied learning yielded significant learning gains for chemistry and disease transmission (Johnson-Glenberg et al., 2014) and in the abstract domain of the electric field (Johnson-Glenberg & Megowan-Romanowicz, 2017). Lindgren et al. (2016) also evaluated the effects of embodied interaction on conceptual understanding and learning engagement where their results corroborated that of Johnson-Glenberg's et al. (2014) desktop simulation and embodied learning findings. Specifically, Lindgren et al. (2016) identified that the embodied learning simulation that was designed to teach critical concepts in physics led to positive results in terms of students' learning gains, engagement, and attitudes towards science.

2.2. Multimodal learning in XR

Learning is often multimodal (Jewitt, 2006) and associated with a variety of modes of communication (Ochao, 2017). Being able to capture the change of mode is critical to the understanding of learning processes. Multimodality contributes to learning via both multimodal instruction and complex multimodal representations created by learners. Multimodal instruction facilitates learning with effectively integrated representations of content across different sensory modalities (e.g., Birchfield et al., 2008).

Various methods have been developed by integrating multimodal data sources and have shown promise to further understanding of learning in an embodied environment. Traditional data such as observation, audio/video recording, and student and instructor discourse can be used to investigate embodied learning. However, such analytical processing of these data has limitations including being time consuming, error-prone, and having limited scalability (Prieto et al., 2018). Advanced technologies now enable the collection of a larger spectrum of multimodal data sources that reflect students' embodied experiences and further inform multimodal learning and instruction. The overall effectiveness of body-based learning activities has been demonstrated, however, there has been less attention given to the progression of embodied actions (e.g., gestures) performed within such a learning environment and the ways that these progressions may be conducive to learning (e.g., Smith et al., 2016). As the availability of multimodal data collected from embodied learning environments increases, identifying analytical approaches that allow for investigation and interpretation of these data becomes imperative.

2.3. Time series clustering analyses

Given the large quantities and ever-increasing complexity of data available, the need for scalable, time-series based analyses are critical (Lin et al., 2012). Scaling and performance necessitate additional revolutions in time series-driven pattern recognition. Many techniques, specifically in clustering, have evolved to meet this challenge. For instance, the KmL (K-Means for Longitudinal data) algorithm was developed out of K-means for longitudinal data (Genolini & Falissard, 2011). A benefit to using this over traditional K-means is that KmL can handle missing data seen frequently in time. Another time series clustering method, Dynamic Time Warping (DTW) clustering, has shown benefits. DTW calculates the minimum differences between two time series that can differ in length and amplitude based on the creation of an optimal warping path to assess similarities through one-to-many mapping (Li et al., 2010). The output of DTW results in a (dis)similarity metric that can then be leveraged in a clustering algorithm. For instance, Mezari and Maglogiannis (2017) used DTW on motion data to recognize gestures. Shen and Chi (2017) also applied DTW by using 36 variables dealing with autonomy (i.e., hitcounts), temporal information (i.e., average times), and actions extracted from student interactions with a tutoring system. They explored different types of clustering methods, suggesting that popular clustering techniques such as K-Means do not account for the differing sequential nature of many educational-based problems. They explored DTW compared to the Euclidean distance and found DTW a viable method for clustering data where participants have differing lengths of time and numbers of interactions. The powerful applications of DTW are often lauded for the ability to evaluate time series of differing lengths and scalability to large quantities of data for pattern recognition.

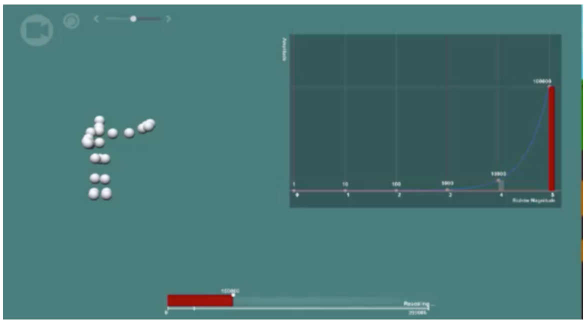
There is a paucity of research in investigating the analysis of fine-grained multimodal interaction data collected in a XR embodied learning environments, and how to interpret the interaction data of learners' embodied actions. The goal of this paper, therefore, is to investigate the massive and dynamic gesture data generated as students interact with an XR embodied science simulation and how such gesture data relate to students' embodied learning experiences. The main contribution of this paper is that we employ a novel analytical method (i.e., DTW clustering) to the fine-grained time-series gesture data. In particular, we adapt the data aggregation methods in Shen and Chi (2017) to investigate different granularities of gesture data and describe how each data granularity reveals different meanings. The present study is an exploratory one, in which we explore and describe different levels of analyses to discover the best data windowing techniques for identifying meaningful sequential patterns of students' embodied actions. In addition, we explore different features derived from gesture data, and we identify potential features for personalized feedback that facilitate students' productive whole-body movements and their learning in the future implementation of XR embodied learning environments.

Our primary hypothesis is that certain patterns identified from students' gesture data (e.g., time spent, volume, or speed on a specific gesture) will be related to their learning performance. For example, students who exhibit an increasing trend in their time spent on gestures that are misaligned with the target mathematical relationship (e.g., doing an addition gesture when a multiplication gesture is called for) to solve a problem will show lower learning gains. We aim to find new ways to analyze multimodal interaction data that will reveal any underlying connections between body movement and learning. This exploratory work lays the groundwork for automated detection of student embodied behaviors that ultimately supports personalized learning.

3. Methods

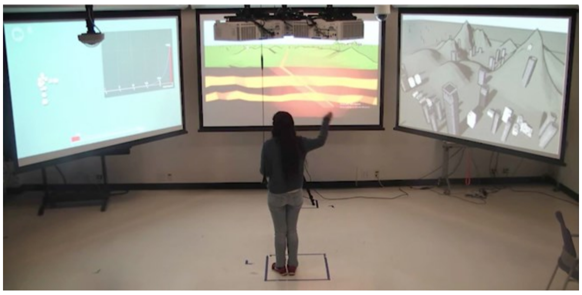
3.1. ELASTIC^{3}S

ELASTIC^{3}S (see Figure 1) was developed using the Unity engine for Microsoft Kinect V2 to empower high school and undergraduate students to build scientific knowledge around the crosscutting concept of scale, proportion, and quantity. This crosscutting concept was used to bridge the science topics of (1) earthquakes and (2) acidity/basicity. To detect students' movements and provide them with real-time skeletal-motion feedback (see Figure 1a), a gesture recognition system was developed using a hierarchical hidden Markov model (described in Junokas et al., 2018). The system is adaptive; that is, the system learns each participant's different types of gestures associated with four mathematical operations (i.e., +1, −1, ×10, and ÷10), which the system uses to recognize the participant's real-time skeletal data. This paper focuses on the gesture-based data that was collected during the earthquake simulation of ELASTIC^{3}S, which explored the concept of linear and non-linear growth through the application of the Richter scale. Students began by developing personal gestures using the metaphorical framing for each mathematical function we identified during early pilot interviews (Alameh et al., 2016). For example, participants were prompted to think of (1) addition as stacking a cube on top of a pile of other cubes, (2) subtraction as kicking one cube out of a pile, (3) multiplication as folding copies of a certain quantity on top of each other, and (4) division as splitting a stack into smaller groups. While students were cued to create gestures that adhered to this framing, the "one-shot" gesture recognition system (Junokas et al., 2018) meant that each student could develop their own personal gestural representation. The gestures that students created allowed them to explore the exponential concepts in different science topics (i.e., earthquakes and acidity/basicity).



(a) Student's Skeleton and Two Bar Graphs

(b) Earth's Fault Line



(c) Three-Screen Simulation Space

Figure 1. Screenshots of the ELASTIC^{3}S XR Earthquake Simulation. *Note.* (a) Two bar graphs indicate the student's input magnitude; (b) The Earth's fault line shows the building of pressure associated with the magnitude; (c) A student is making a multiplication gesture.

Once students completed the training phase, students were provided five sequential tasks that varied in difficulty. First, they are prompted to set the amplitude of the seismic waves to create a different magnitude of earthquake. Students begin with a straightforward task, which we call M2, where they create a magnitude 2 earthquake (corresponds to 100 amplitude units, 10^2). The most efficient use of gestures to complete this task would be, starting at 0, to add 1, then multiply by 10 and multiply by another 10, resulting in an amplitude of 100 units.

Each task moved students through varying complexities of their gestures as well as conceptual understandings; for example, M3.5 requires students to apply their acquired knowledge of the exponential nature of the Richter scale in order to create a magnitude of 3.5 earthquake. This magnitude is not halfway between amplitudes of magnitude 3 and 4 and requires an in depth understanding of exponential growth. The task names, the corresponding amplitude, and level of difficulty are listed in Table 1.

Table 1. Simulation tasks and difficulty

Task order	Task name	Corresponding amplitude	Difficulty level*
1	M2	$10^2 = 100$	Mid - Hard
2	M3	$10^3 = 1{,}000$	Easy - Mid
3	M3.5	$10^{3.5} = 3{,}162.28$	Hard
4	M7	$10^7 = 10{,}000{,}000$	Easy
5	M8	$10^8 = 100{,}000{,}000$	Easy

Note. *Relative to other given tasks.

3.2. Participants

A total of twenty-four undergraduate students from the midwestern region of the United States participated in the earthquake simulation, which consisted of a pre-test, simulation session, and post-test. Following the IRB-approved protocol, each participant individually completed a task-based interview consisting of (1) a pre-test, (2) a simulation session lasting about 30 minutes, and (3) a post-test in a lab with a facilitator in the room. The pre-/post-tests consisted of questions that assessed their understanding of earthquakes and linear/non-linear growth in both the context of earthquakes and new contexts (see details in 3.3.3). During the simulation session, the participants were asked to engage with the earthquake simulation. Both the pre-/post-tests and simulation session were audio and video recorded. Unfortunately, the logs of early student participants were not recorded, with one additional student identified as an outlier via visualizations (see section 3.3.2) thus leaving twelve students' data available for this study. Of these twelve students (mean age = 19.2), 58% (n = 7) stated they were female, 66% were white, two preferred not to answer, one student was Asian, and one identified as multiracial.

3.3. Data sources

3.3.1. Kinect data

The participants completed five tasks in a 30-minute earthquake simulation session. This paper focuses on the data collected via Microsoft Kinect V2 during the simulation session. Datapoints were collected at a rate of approximately 30 frames per second, in which the 3D coordinates (x, y, z) of 25 skeleton joints relative to the Kinect were tracked and recorded. Particularly, each joint's coordinates were recorded based on positions relative to the position of the Kinect sensor (see Figure 2).

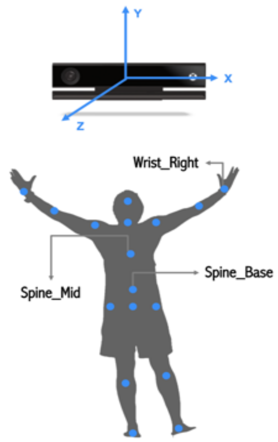


Figure 2. Visual representations of data collection including joints of interest

3.3.2. Data preprocessing and feature extraction

The first level of preprocessing involved data cleaning through annotation of session recorded video, documenting when participants started and stopped each gesture and the type of gesture being performed. This information was then synced with the Kinect data and resulted in a dataset that contained all recorded data, the gesture type, gesture order, timestamp, and the task being completed. We excluded any Kinect data without annotations. We also visualized the data to remove unintended movements recorded by the Kinect. It was identified that one participant had extreme outliers in M2 and M3.5 in both volume and speed, which appeared to be due to system glitches. That student was therefore removed from the dataset.

Using the cleaned data, we developed various simulation use measures (see Table 2). First, the *speed* of a gesture was measured by finding the magnitude of the velocity of a given joint position, formally expressed as (p_n: n_{th} joint position, t: time of a given joint position, s: speed):

$$\Delta p = p_1 - p_0$$
$$\Delta t = t_1 - t_0$$
$$\Delta s = \left|\frac{\Delta p}{\Delta t}\right|$$

This can be ultimately measured at individual, combinations, and/or the complete joint positions. The *volume* of a gesture was measured by finding the product of the Euclidean distance between the maximum and minimum points of given joint positions at each dimension (*x, y, z*), formally expressed as (v: volume):

$$d_x = \sqrt{(x_{max} - x_{min})^2}$$
$$d_y = \sqrt{(y_{max} - y_{min})^2}$$
$$d_z = \sqrt{(z_{max} - z_{min})^2}$$
$$V = d_X * d_Y * d_Z$$

This rectangular projection of volume was measured on all joints, providing a spatial perspective to a performed gesture. We then used the distance between two joint positions: spine_base and spine_mid (see Figure 2; these data points represent each participant's height) for the normalization, rescaled by each participant's body size.

While such frame-level data (referred to as "frame granularity") showed in-moment data, we wished to identify what level of aggregation was most important and made the most sense. This was completed by comparing multiple aggregations on frame granularity data (Shen & Chi, 2017). We therefore aggregated the frame granularity data by each gesture (referred to as "gesture granularity") to see if there would be a difference between our two types of granularity: gesture and frame. We found that compared to the frame granularity, the gesture granularity was more interpretable, as we were interested in the gesture as a whole, rather than a single frame. As for speed and volume of each gesture, the maximum, minimum, variance, and average were created for both variables on each distinct gesture to retain as much information as possible.

Further, we needed to know how long the gesture took (i.e., timestamp at the last frame – timestamp at the first frame) and the type of gesture that was made. We named these measures as: "total time spent on" addition, subtraction, multiplication, and division. In addition, during the aggregation process, we naturally lost the granularity of frames. We therefore included how many "frames" were made in the creation of each gesture, which was named as "number of frames." Table 2 shows the final list of variables (i.e., simulation use measures), the result of the aforementioned data preprocessing and feature extraction efforts.

Table 2. Simulation use measures for gesture level analysis

Variable name	Description	Data type	Example
Task	Earthquake simulation tasks	Character	M2
Volume	Four features (average, variance, max, min) generated from the aggregation of [1]normalized volume.	Decimals (m^3)	0.0033
Speed	Four features (average, variance, max, min) generated from the aggregation of [1]normalized speed.	Decimals (meter/second)	0.0437
Number of frames	Number of rows condensed to produce aggregated row at the gesture level	Whole Number	79
Time spent (on gesture)	Each gesture was considered a feature, resulting in four features (addition, subtraction, multiplication, division).	Decimals (second)	1.2853

Note. [1]Normalization occurred on the frame level which standardized each motion across all participants for comparison.

3.3.3. Learning performance data

The four researchers scored each participant's understanding of earthquake concepts and exponential growth during pre- and post-tests (Kang et al., 2018). Table 3 shows the descriptive statistics of learning performance data. During pre-/post-tests, the facilitator asked several questions to see the participant's understanding of earthquakes (i.e., conceptual knowledge) and exponential growth in the earthquakes (i.e., exponential knowledge) and new contexts (i.e., transfer knowledge) (See sample questions in Appendix A). In this study, the normalized gain score of each category (conceptual, exponential, and transfer knowledge) was calculated by ((posttest score) – (pretest score)) / ((total available score) – (pretest score)) (Hake, 1998). These learning performance measures were used later to examine if any identified clusters showed significant intergroup differences of learning performance.

Table 3. Descriptive statistics for learning performance

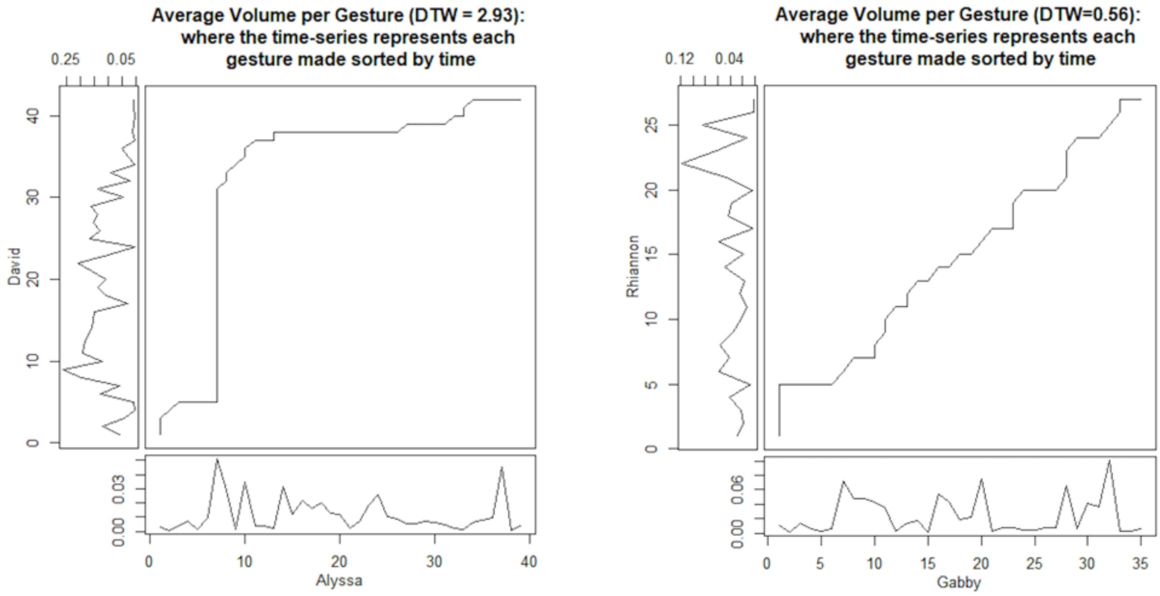
	Conceptual			Exponential			Transfer		
	Pre	Post	[1]Gain	Pre	Post	[1]Gain	Pre	Post	[1]Gain
Mean	3.08	8.25	0.75	4.63	7.67	0.28	3.00	2.88	-0.08
SD	0.90	1.08	0.15	3.48	2.63	0.28	1.65	1.86	0.50

Note. [1]Normalized gain score (Hake, 1998).

3.4. Analyses

The gesture granularity data utilized in this study are multivariate time series (see the variables in Table 2). Several considerations were observed to maximize effectiveness of this multivariate information in categorizing individuals based on their embodied learning. For example, time series data is a critical aspect of understanding student embodied learning process, and individuals did not complete a task in the same number of frames or gestures. DTW takes into consideration time series of different lengths, resulting in a computed similarity value of each pair of time series data, which is similar to a Euclidean distance (Shen & Chi, 2017). This results in a distance matrix comparing distance similarity values with each other.

For a single variate, visual example on warping paths and the generation of a similarity metric we present two examples in Figure 3: (1) DTW dissimilarity = 2.93 between David and Alyssa (Figure 3a), (2) DTW dissimilarity = 0.56 between Gabby and Rhiannon (Figure 3b). In each graph, we see three panels in each warping path. Each student's behavior over the course of time for a given variable is shown in each vertical (left) and horizontal (lower) panel. The central panel is the cost matrix, where we see the warping path between the two students' average volume patterns over time. A score of zero, or a diagonal line, indicates that the patterns are identical and no warping was required to "match" one pattern to another. The more warping that is required, the larger the metric is, indicating more dissimilarity between two students' behaviors.



(a) Comparison of David and Alyssa, DTW = 2.93 (b) Comparison of Gabby and Rhiannon, DTW = 0.56

Figure 3. Average volume comparison between a pair of students

Further, we sought to understand how different levels of the data could yield meaningful clustering results that address our objectives. Three levels: (1) task level, (2) subsequence level, and (3) all tasks level were created to view the data in multiple ways (see Table 4). For example, M3.5 on the task level used gesture data from only task 3.5. For M2/3/3.5 in subsequence level, we included the student's first gesture made in M2 to their last in M3.5. Therefore, subsequence level was ultimately used to understand the overall trajectory of the behavior up to that point. Task level and subsequence level clustered twelve matrices, each of which included each of twelve participants' multivariate gesture granularity data for each level of analysis. Finally, for all tasks level, we included all participants' gesture granularity wherein each student's behavior per task was evaluated against all other students' task behavior. That is, all task level clustered a total number of sixty matrices (i.e., twelve students × five tasks = sixty matrices), each of which included the multivariate gesture granularity data each participant made during each session.

Once the DTW value or metric was derived for each pair of data, the values were used in the clustering process. We selected the Hierarchical clustering method using Wards Linkage to ultimately partition students into clusters. The optimal number of clusters was identified using Silhouette method for each analysis level: 3 clusters. This was all completed in the tsclust package in R. To examine statistically significant differences of each variable across three clusters in each level of analysis, we first checked the assumptions of one-way

ANOVA. Along with our small sample size, the normality and homogeneity of variances were violated. Therefore, we performed Kruskal-Wallis non-parametric analyses of gesture and learning performance variables (i.e., normalized gain on conceptual, exponential, and transfer knowledge) of each cluster generated from task level, subsequence level, and all tasks level. Then, we conducted post hoc tests with the Bonferroni adjustment (Kruskal & Wallis, 1952) to examine statistically significant differences of each variable between each pair of three clusters.

Table 4. Different levels considered for clustering and other analyses

Analysis level	Tasks included in each analysis (# gesture granularity data)	Data structure examples[1]
Task level	M2 (n = 64)	A.M2, B.M2 … T.M2
	M3 (n = 21)	A.M3, B.M3 … T.M3
	M3.5 (n = 173)	A.M3.5, B.M3.5… T.M3.5
	M7 (n = 86)	A.M7, B.M7… T.M7
	M8 (n = 21)	A.M8, B.M8 … T.M8
Subsequence level	M2 (n = 64)	A.M2, B.M2 … T.M2
	M2, M3 (n = 85)	(A.M2+A.M3), (B.M2+B.M3) … (T.M2+T.M3)
	M2, M3, M3.5 (n = 258)	(A.M2+A.M3+A.M3.5) … (T.M2+T.M3+T.M3.5)
	M2, M3, M3.5, M7 (n = 344)	(A.M2+A.M3+A.M3.5+A.M7) …
	M2, M3, M3.5, M7, M8 (n = 365)	(A.M2+A.M3+A.M3.5+A.M7+A.M8) …
All tasks level	M2, M3, M3.5, M7, M8 (n = 365)	A.M2, B.M2 … T.M7, T.M8

Note. [1]Letters that come prior to task information are shortened aliases for students. A.M2 is Alyssa's data from task M2 exclusive.

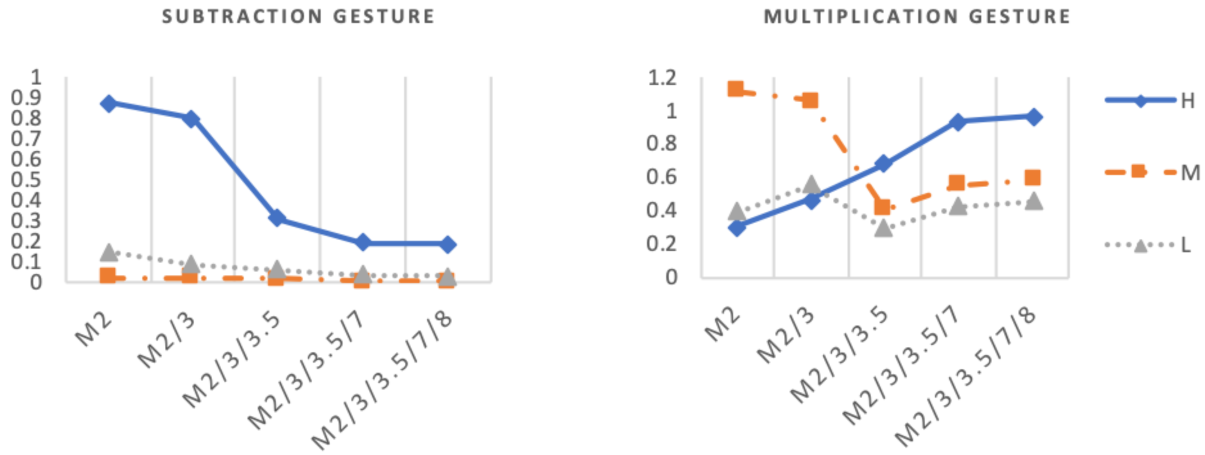
4. Findings

4.1. Task level vs. Subsequence level

To identify students' gestural patterns, we first explored two different levels: task level and subsequence level. As described in 3.4, non-parametric analyses examined statistically significant differences of gesture and learning variables across three clusters. Of interest, the number of frames variable showed the significant intergroup differences in every task and subsequence level. Overall gesture characteristics of each cluster were identified as "High-Frame," "Mid-Frame," and "Low-Frame" mainly using each cluster's mean rank of number of frames. Table 5 shows the characteristics of each cluster in the subsequence level analyses. In particular, we used the results of K-W analyses and post hoc tests by including selected statistically significant variables (i.e., simulation use measures, learning performance variables) identified in each segment analysis.

The task level analyses showed the participants' significant gestural behavior differences across three clusters during each task. One interesting finding is that a High-Frame cluster showed the lowest transfer change ($High_{\_mean\ rank}$ = 2.0, $Mid_{\_mean\ rank}$ = 8.0, $Low_{\_mean\ rank}$ = 8.0) during the M7 task ($\chi^2(2)$ = 7.748, p < .05) and the lowest exponential change ($High_{\_mean\ rank}$ = 4.20, $Mid_{\_mean\ rank}$ = 9.50, $Low_{\_mean\ rank}$ = 4.75) during the M8 task ($\chi^2(2)$ = 6.009, p < .05). The High-Frame clusters during M7 and M8 show the same tendency of longer time spent on gesture multiplication (M7: $\chi^2(2)$ = 55.658, p = .000; M8: $\chi^2(2)$ =10.515, p = .005). Potentially at this point, the students who showed this gestural characteristic could be provided better scaffolding, which could, in turn, be more successfully transferred to novel contexts.

The subsequence level analyses also identified three clusters in each aggregated segment. As shown in Table 5, within each subsequence segment, all High-Frame clusters show the longest time spent on gesture subtraction, while all Mid-Frame clusters show the shortest time spent on gesture subtraction, as evaluated using K-W non-parametric testing. The two gestures, subtraction and division, indicate a student made an adjustment of the gestures they have added, realizing previous gestures were made incorrectly. To complete the tasks in the optimal way, these two gestures are not required at all. The presence of these patterns in the data suggests that students may have had difficulty employing the correct sequence of gestures that were required. Figure 4 shows the average time spent trend (i.e., the average time spent per a single gesture) on two gesture types: gesture multiplication and gesture subtraction. One interpretation may be that students initially had trouble grasping the system, as the average time spent on Gesture Subtraction decreases over the segments (see Figure 4a). However, we also recognize that the High-Frame groups show significantly more time spent on gesture subtraction during the first two tasks than the other groups which may signify the overall difficulty these students had and may require additional support in situ.



(a) Average Time Spent Cluster Trends on Gesture Subtraction

(b) Average Time Spent Cluster Trends on Gesture Multiplication

Figure 4. Average time spent trends on gesture

As shown in Figure 4b, High-Frame clusters show an increasing trend, while the Mid- and Low-Frame clusters show a decreasing trend. Particularly, the High-Frame cluster show the lowest time spent on multiplication gestures for the first two subsequence segments: M2 and M2/3.5. The High-Frame clusters then overtook the other two clusters when the later tasks' data (i.e., M3.5, M7) were added. This increasing average time spent on multiplication gesture may indicate this is a point of tension for the student. During the last two segments, compared to the Mid-Frame clusters, the students in the High-Frame clusters also appeared to achieve lower normalized gain on transfer knowledge, indicating they were unable to transfer their knowledge to new contexts.

In Figure 5, the normalized volume per gesture shows the similar trends across all clusters; that is, the increasing trend on the M2/3 and M2/3/3.5/7/8 subsequences. When we included M3 in the subsequence (i.e., the M2/3 segment), the average volume made for each gesture increased. This was echoed when M8 data was included in the full sequence. The High-Frame clusters showed consistently the lowest average of volume over the entire task progression. Task M8 (from M7) is the easiest task that requires students to go over the similar thinking process they practiced during M2 and M3. Additionally, the Mid-Frame group showed the sharpest increasing trend of average volume from M7 to M8. Given the highest transfer knowledge gain Mid-Frame groups showed, the tendency of increasing volume during each following task of M2 or M7 may indicate students' confidence on their gestures.

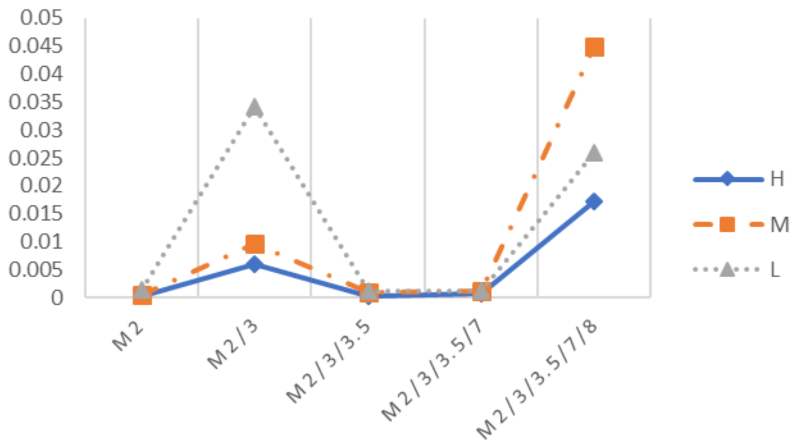


Figure 5. Average normalized volume trend

Table 5. Cluster characteristics: Gesture and learning - subsequence level

Tasks included	# Granularity data[1]	Selected key variables[2]	High-Frame Cluster	Mid-Frame Cluster	Low-Frame Cluster	Post hoc results[5]
M2	Total = 64 High = 20 Mid = 21 Low = 20	Frames ($\chi^2(2) = 25.121$, $p = .000^{**}$)	MR[4] = 45.33	MR = 36.55	MR = 17.65	**High-Low; **Mid-Low
		Timespent_S[3] ($\chi^2(2) = 18.017$, $p = .000^{**}$)	MR = 43.25	MR = 26.19	MR = 28.91	**High-Mid; **High-Low
		Timespent_M ($\chi^2(2) = 17.228$, $p = .000^{**}$)	MR = 23.55	MR = 44.95	MR = 28.91	**High-Mid; **Mid-Low
M2/ M3	Total = 85 High = 25 Mid = 30 Low = 30	Frames ($\chi^2(2) = 25.121$, $p = .000^{**}$)	MR=63.73	MR = 36.55	MR = 17.65	**High-Low; **Mid-Low
		Timespent_S[3] ($\chi^2(2) = 18.017$, $p = .000^{**}$)	MR = 43.25	MR = 26.19	MR = 28.91	**High-Mid; **High-Low
		Timespent_M ($\chi^2(2) = 17.228$, $p = .000^{**}$)	MR = 23.55	MR = 44.95	MR = 28.91	**High-Mid; **Mid-Low
M2/ M3/ M3.5	Total = 258 High = 57 Mid = 147 Low = 52	Frames ($\chi^2(2) = 107.164$, $p = .000^{**}$)	MR = 202.11	MR = 129.02	MR = 54.13	**High-Mid; **High-Low
		Volume_Average ($\chi^2(2) = 17.482$, $p = .000^{**}$)	MR = 106.7	MR = 146.11	MR = 106.88	**High-Mid; **Mid-Low
		Speed_Average ($\chi^2(2) = 10.973$, $p = .004^{**}$)	MR = 131.82	MR = 139.11	MR = 99.42	**Mid-Low
		Timespent_S ($\chi^2(2) = 20.970$, $p = .000^{**}$)	MR = 146.32	MR = 123.48	MR = 128.33	**High-Mid; *High-Low
M2/ M3/ M3.5/ M7	Total = 344 High = 98 Mid = 138 Low = 108	Frames ($\chi^2(2) = 177.092$, $p = .000^{**}$)	MR = 268.41	MR = 173.74	MR = 83.88	**High-Mid; **Mid-Low; **High-Low
		Volume_Average ($\chi^2(2) = 10.113$, $p = .006^{**}$)	MR = 155.46	MR = 193.13	MR = 161.6	*High-Mid; *Mid-Low
		Speed_Average ($\chi^2(2) = 9.315$, $p = .010^{*}$)	MR=183.58	MR=183.4	MR = 148.52	*Mid-Low; *High-Low
		Timespent_S ($\chi^2(2) = 17.125$, $p = .000^{**}$)	MR = 185.27	MR = 165.19	MR = 170.26	**High-Mid; *High-Low
		Timespent_M ($\chi^2(2) = 15.129$, $p = .001^{**}$)	MR = 202.1	MR = 167.29	MR = 152.3	**High-Mid; *High-Low
		Transfer Change ($\chi^2(2) = 8.437$, $p = .015^{*}$)	MR = 2.00	MR = 8.80	MR = 7.00	*High-Mid
M2/ M3/ M3.5/ M7/ M8	Total = 365 High = 101 Mid = 145 Low = 119	Frames ($\chi^2(2) = 11.192$, $p = .000^{**}$)	MR = 287.16	MR = 187.42	MR = 89.21	**High-Mid; **Mid-Low; **High-Low
		Volume_Average ($\chi^2(2) = 7.660$, $p =.022^{*}$)	MR = 165.39	MR = 201.16	MR = 175.82	*High-Mid
		Timespent_S ($\chi^2(2) = 17.874$, p	MR = 196.38	MR = 175.7	MR = 180.53	**High-Mid; **High-Low

	=.000[**])				
	Timespent_M ($\chi^2(2)$ = 16.575, p =.000[**])	MR = 215.51	MR = 178.53	MR = 160.84	[*]High-Mid; [**]High-Low
	Transfer Change ($\chi^2(2)$ = 8.437, p =.015[*])	MR = 2.00	MR = 8.80	MR = 7.00	[*]High-Mid

Note. [1]Number of gesture granularity data included in the analysis. [2]Key selected significant variables from non-parametric analyses of each cluster. [3]Total time spent on gesture (A: Addition, S: Subtraction, M: Multiplication, D: Division). [4]Mean Rank. [5] Post hoc test results using Bonferroni correction. [*]indicates p-value < .05. [**]indicates p-value < .01.

4.2. All tasks: Membership transfer

To examine how each student shifted cluster memberships over five different tasks, we completed another analysis in which clustering took place where each task was evaluated for each person. Three clusters were selected as the optimal number of clusters. Similarly, the number of frame feature shows statistically significant differences across three clusters; therefore, we also used the mean ranks of number of frames to label each cluster. The non-parametric tests and post hoc analyses showed some interesting statistically significant variables across the three clusters: time spent on gesture addition ($\chi^2(2)$ = 38.597, p = .000) and time spent on gesture subtraction ($\chi^2(2)$ = 7.020, p = .030). For example, High-Frame cluster shows the longest time spent on gesture addition tendency ($High\_{mean\ rank}$ = 212.95), while Mid-Frame cluster shows the shortest time spent tendency ($Mid\_{mean\ rank}$ = 114.5). Such patterns of High-Frame cluster may indicate students were less strategic or struggling, since the use of gesture addition is indicative of less understanding of non-linear relationships between the amplitudes of seismic waves and the magnitude.

Table 6. Membership transfer at all tasks level

Pseudonym	M2	M3	M3.5	M7	M8	[2]Key learning gains
Alyssa	H[1]	H	H	H	M	Low Exponential/Conceptual/Transfer
Blair	H	H	H	H	M	Low Transfer
David	H	M	H	L	M	None
Gabby	H	M	H	L	L	Low Conceptual; High Exponential
George	L	L	L	L	M	None
Louise	L	L	L	L	L	High Exponential/Transfer
Matthew	L	M	L	L	M	Low Exponential
Mindy	L	M	L	L	M	None
Rhiannon	L	M	H	L	L	High Exponential/Transfer
Rosalind	L	M	L	L	M	High Exponential
Steven	M	M	H	H	M	Low Conceptual/Exponential/Transfer
Tabitha	M	M	L	L	M	Low Conceptual; High Transfer

Note. [1]This is an abbreviation of each cluster (H: High-Frame, M: Mid-Frame, L: Low-Frame). [2]This is a summary of each student's learning gain in each category of pre-/post-test scores. Low indicates the 25% of participants who had the least learning gains, where High represents the 25% participants who made greater learning gains.

Table 6 shows each participant's clustering membership transfer over the five tasks and their key learning gain. Notably, the majority of students were assigned to a High-Frame cluster during M3.5, the most challenging task. After the completion of the first two tasks (M2, M3), we expect students to no longer stay in High-Frame cluster during the second set of the similar practice (M7, M8), where they were able to apply what they learned from the earlier tasks. Therefore, the students who exhibit the patterns staying in High-Frame cluster (i.e., Alyssa, Blair) or switching to High-Frame cluster (i.e., Steven) seemed unable to figure out how to use the system until the end of the simulation session. These students' lower learning performance indicated that such patterns may be an indicator of struggles. The participants who tended to stay in Low-Frame cluster toward the end of the simulation showed better learning gains in both exponential and transfer knowledge (i.e., Louise, Rhiannon). Understanding such cluster membership transfer reveals the trend of each student's gesture use across five different tasks. This may suggest ideal gesture behavior that leads to a positive learning outcome, which needs further research to verify the relationship.

5. Discussion

Recent research has highlighted students' challenges in understanding complex and abstract STEM concepts and the role that embodiment can play in overcoming these challenges (e.g., Duijzer et al., 2019; Stieff et al., 2016). ELASTIC^{3}S is an XR environment that was designed to facilitate crosscutting concept knowledge-building (scale, proportion, and quantity) by having students develop gestures to express ideas about linear/non-linear growth in different science domains. In this study, we were particularly interested in understanding embodied interaction data collected from an immersive XR platform as students were engaged in learning about non-linear relationships in one specific science domain (i.e., earthquakes). This study employed the DTW clustering method to explore and better understand students' embodied learning processes based on the fine-grained time-series gesture data recorded from the Microsoft Kinect.

It is important to explore different data granularities as they can reveal meaningful patterns in different contexts (e.g., Shen & Chi, 2017). To best describe the embodied learning experiences in this simulation, we first explored different types of data granularity: frame vs. gesture. The frame granularity did not tell us as much information as we were interested in the gesture as a whole, rather than a single frame; therefore, the gesture granularity yielded more interpretable results. We further explored different levels of analyses: (1) task, (2) subsequence, and (3) all tasks, to discover the best data windowing techniques for sequential patterns of embodied learning. This highlights the importance of data exploration to best tell the story in a certain context, and contributes on the literature of understanding embodied learning process within technology-enhanced learning environments (e.g., Price et al., 2016), especially by applying different analytical applications of fine-grained time-series embodied interaction data.

In ELASTIC^{3}S, students complete a series of five tasks with the goal of reaching a specific value on the Richter scale. Moving through the tasks, both students' gesture use and their conceptual understandings are advanced as they become more familiar with the system. The subsequence level analyses revealed more meaningful temporal patterns than task level analyses, as their captured behavior in a given task was placed in context with previous tasks. Further, any changes that occurred at each segment were certainly influenced by performance the time before.

The results indicate that certain gestural interaction trends of students' simulation use identified in Mid-Frame clusters (e.g., a decreasing trend of time spent on gesture multiplication, relatively lower time spent on gesture subtraction over the entire segments, relatively higher average volume over the last three segments) seemed to be associated with higher learning gains. This suggests their increasing trend of volume toward the later segments may indicate the students' higher confidence of using gesture as they become more familiar with the system and the exponential growth concepts, which needs further research to verify such relationship. The characteristics of a High-Frame cluster including the relationship with the learning gain measures showed that students exhibiting such characteristics (e.g., an increasing trend of time spent on gesture multiplication, relatively lower average volume over the last three segments) seemed less likely to learn the core ideas presented in the simulation. The results of the M2/3/3.5/7 segment, showed significant transfer knowledge gain differences between clusters. This may be an indicator of the needs of guidance for some students who showed the characteristics of High-Frame cluster based on their simulation use up to M7. For example, the system can track each individual's subsequence data aggregated from M2 to M7. If a student shows relatively high speed, longer time spent on gesture subtraction or gesture multiplication, or smaller volume, the system may provide certain prompts so that students can reflect on what they have done up to the current task. In this way, students will be able to receive more practice. This is aligned with the literature that highlights the importance of capturing the subtle change of students' multimodal interaction to understand their learning process and further yield positive learning performance (Jewitt, 2006; Ochoa, 2017). It is noteworthy that our sample size is relatively small (n=12) for making predictions that generalize to other groups of students. A larger sample is needed to validate that such patterns are representative of the behaviors of a larger group.

The subsequence level analyses overall tracked different gestural interaction patterns during each task within the context of past tasks and gestures, suggesting which task the participants may have struggled in completing. We further examined each individual student's cluster membership transfer (all tasks level) indicating an individual's sequential pattern of simulation use. The students who switched to Low-Frame cluster toward to the last task tended to achieve more learning. However, the students who exhibited the tendency to backtrack (i.e., the longest time spent on gesture subtraction) and struggled with understanding non-linear contexts (i.e., the longest time spent on gesture addition) showed lower conceptual gain. This suggests the clustering transfer trend might be an indicator of more support or practice being needed. These findings are descriptive in nature, and require further investigation for causal inferences between gestural patterns and learning performance.

Research (e.g., Abrahamson & Lindgren, 2014; DeSutter & Stieff, 2017) have identified components (e.g., activities, materials, and facilitation) of embodied learning environments and design principles that can be applied to the learning environment design. In particular, facilitation should be an integral part of the embodied learning environment, in which learning is facilitated by situated and timely feedback. Since body cueing is one of the ways embodied interactions are prompted around learning content, it is a way to integrate students' understanding of new knowledge with an embodied simulation (Lindgren, 2014). Therefore, it is critical to track students' interactions with the learning environment and to design effective cues that facilitate productive whole-body movements (Black et al., 2012; Lindgren & Johnson-Glenberg, 2013).

Overall, certain temporal patterns using different simulation use measures can be used for early detection of students' struggles throughout the process of exploring embodied activities within an XR learning environment such as ELASTIC^{3}S. Advanced adaptive learning technologies may help to engage the learners with personalized prompts based on kinematic markers to enhance students' cognitive activities in the process of learning. Most of the current studies in the area of embodied learning have been conducted in a laboratory environment (e.g., Lindgren & Johnson-Glenberg, 2013), where a researcher or teacher is present throughout the learning activities and provides a participant with just-in-time guidance whenever needed. Exploring potential features driven from multimodal data is critical in the future implementation of embodied learning environment in either a formal or informal setting (e.g., Ochoa, 2017).

6. Limitations and future work

This study has limitations that should be addressed in future research. First, the number of participants is notably small (*n* = 12). However, the Kinect logfiles contain a massive gesture dataset for each participant. This yielded the adequate amount of data (see Table 4) that represents valid and reliable behaviors of the participants, which allowed for the analyses we conducted in this study. In future work, we plan to include a larger sample to validate (1) whether the captured patterns are representative of the behaviors of a larger group and (2) the relationships between the patterns and learning performance. Second, it should be noted that we did not collect the participants' academic background such as fields of study or previous experiences in STEM, while we recruited the participants from a general educational psychology course where the majority of students' fields of study was not STEM.

Third, we identified the cluster patterns including volume and speed features, which suggested their affective states such as confidence of using gesture rather than cognitive states. This needs further research to verify such a relationship. This also suggests interesting lines of follow-up inquiry on the causal relationship between kinematic features or gesture pattern trajectories and affective or cognitive states. Lastly, we extracted the simulation use variables by considering all joint information in one frame. There is a lot more nuance to 3-dimensional metrics especially in spatial positions. Therefore, future studies should further explore other features, such as focusing on one joint for all frames or all joints and all frames, as each new feature will return different metrics and representations. We hope the volume features we used in this study are a good starting point for further exploration. The analytical approach used in this study indicates the potential of kinematic features as key indicators of the quality of learner perceptions and comprehension, and the potential need for gestural interaction guidance, which can further support their learning in other domains.

7. Conclusion

Previous studies have shown that the embodied learning simulation, ELASTIC^{3}S, has an overall a positive impact on students' understanding of content objectives and the crosscutting concept of non-linear growth. Those studies highlighted the critical features of embodied simulations that facilitate student reasoning. Technological advances in gesture recognition allow for the creation of XR environments that can track and respond to students' gestures in real time. This study therefore investigated how the gestures learners perform in these environments relate to their subsequent learning, and how understanding the features of productive gestures can be applied to future embodied XR learning environment design. We applied multivariate Dynamic Time Warping for clustering to identify gestural patterns in ELASTIC^{3}S as evidence for understanding learning processes. Our findings showed that identified patterns of simulation use were indicative of students' comprehension and struggles with learning target ideas.

The main contribution of this paper is to apply an underutilized analytical approach to understand students' gestural interactions with embodied XR learning environments. Specifically, this study contributes to the early work of detecting sequential gesture patterns that represent students' embodied learning experiences by applying different data granularities (frame vs. gesture) and different levels of analysis (task, subsequence, and all tasks). Different levels of analyses applied in this study highlights the importance of considering various ways of structuring data, which can reveal more meaningful patterns of simulation use and serves as evidence of a more positive embodied learning experience. The results of this study pave the way for further research on the design of XR embodied simulation environments that provide real-time guidance and promote a more powerful and adaptive learning experience.

We proposed potential kinematic features for personalized feedback that may facilitate students' productive whole-body movements and learning. It is worth noting that future research including lager samples is needed to validate the present findings. We believe such analytical applications explored in this paper provide guidance for researchers to replicate or adapt when dealing with fine-grained time-series gesture data within XR embodied environments.

References

Abrahamson, D., & Lindgren, R. (2014). Embodiment and embodied design. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (2nd ed., pp. 358–376). Cambridge, UK: Cambridge University Press.

Alameh, S., Morphew, J., Mathayas, N., & Lindgren, R. (2016). Exploring the relationship between gesture and student reasoning regarding linear and exponential growth. In the *Proceedings of the 12th International Conference of the Learning Sciences* (pp. 1006–1009). Singapore: International Society of the Learning Sciences.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*(1), 617–645. doi:10.1146/annurev.psych.59.103006.093639

Birchfield, D., Thornburg, H., Megowan-Romanowicz, M. C., Hatton, S., Mechtley, B., Dolgov, I., & Burleson, W. (2008). Embodiment, multimodality, and composition: Convergent themes across HCI and education for mixed-reality learning environments. *Advances in Human-Computer Interaction*. doi:10.1155/2008/874563

Black, J., Segal, A., Vitale, J., & Fadjo, C. (2012). Embodied cognition and learning environment design. In D. Jonassen, & S. Land (Eds.), *Theoretical foundations of student-centered learning environments* (2nd ed., pp. 198–223). New York, NY: Routledge.

Crowder, E. M. (1996). Gestures at work in sense-making science talk. *Journal of the Learning Sciences, 5*(3), 173–208.

DeSutter, D., & Stieff, M. (2017). Teaching students to think spatially through embodied actions: Design principles for learning environments in science, technology, engineering, and mathematics. *Cognitive Research: Principles and Implications, 2*(1), 22. doi:10.1186/s41235-016-0039-y

Duijzer, C., Van den Heuvel-Panhuizen, M., Veldhuis, M., Doorman, M., & Leseman, P. (2019). Embodied learning environments for graphing motion: A Systematic literature review. *Educational Psychology Review, 31*(3), 597-629.

Gallagher, S. (2006). *How the body shapes the mind*. New York, NY: Clarendon Press.

Gallagher, S., & Lindgren, R. (2015). Enactive metaphors: Learning through full-body engagement. *Educational Psychology Review, 27*(3), 391–404.

Genolini, C., & Falissard, B. (2011). Kml: A Package to cluster longitudinal data. *Computational Methods Programs Biomed, 104*(3), 112-121.

Glenberg, A. M. (2008). Embodiment for education. In P. Calvo & A. Gomila (Eds.), *Handbook of Cognitive Science* (pp. 355–372). doi:10.1016/B978-0-08-046616-3.00018-9

Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wires Cognitive Science, 1*(4), 586–596.

Goldin-Meadow, S. (2011). Learning through gesture. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(6), 595-607.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A Six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74. doi:10.1119/1.18809

Han, I., & Black, J. B. (2011). Incorporating haptic feedback in simulation for learning physics. *Computers & Education, 57*(4), 2281–2290.

Jewitt, C. (2006). *Technology, literacy and learning: A Multimodal approach*. New York, NY: Routledge.

Johnson-Glenberg, M. C., Birchfield, D., Savvides, P., & Megowan-Romanowicz, C. (2011). Semi-virtual embodied learning-real world stem assessment. In L. Annetta & S. C. Bronack (Eds.), *Serious educational game assessment: Practical* *methods and models for educational games, simulations and virtual worlds* (pp. 241–257). Rotterdam, The Netherlands: Sense Publishers.

Johnson-Glenberg, M. C., Birchfield, D. A., Tolentino, L., & Koziupa, T. (2014). Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. *Journal of Educational Psychology, 106*(1), 86–104.

Johnson-Glenberg, M. C., & Megowan-Romanowicz, C. (2017). Embodied science and mixed reality: How gesture and motion capture affect physics education. *Cognitive Research: Principles and Implications, 2*(1), 24. doi:10.1186/s41235-017-0060-9

Junokas, M. J., Lindgren, R., Kang, J., & Morphew, J. W. (2018). Enhancing multimodal learning through personalized gesture recognition. *Journal of Computer Assisted Learning, 34*(4), 350–357.

Kang, J., Lindgren, R., & Planey, J. (2018). Exploring emergent features of student interaction within an embodied science learning simulation. *Multimodal Technologies and Interaction, 2*(3), 39.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association, 47*(260), 583–621.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Li, Y., Chen, H., & Wu, Z. (2010). Dynamic time warping distance method for similarity test of multipoint ground motion field. *Mathematical Problems in Engineering*. doi:10.1155/2010/749517

Lin, J., Williamson, S., Borne, K. D., & DeBarr, D. (2012). Pattern recognition in time series. In *Advances in Machine Learning and Data Mining for Astronomy* (pp. 617–645). New York, NY: CRC Press.

Lindgren, R., & Johnson-Glenberg, M. (2013). Emboldened by embodiment: Six precepts for research on embodied learning and mixed reality. *Educational Researcher*. doi:10.3102/0013189X13511661

Lindgren, R. (2014). Getting into the cue: Embracing technology-facilitated body movements as a starting point for learning. In *Learning Technologies and the Body* (pp. 51–66). doi:10.4324/9781315772639-9

Lindgren, R., Tscholl, M., Wang, S., & Johnson, E. (2016). Enhancing learning and engagement through embodied interaction within a mixed reality simulation. *Computers & Education, 95*, 174–187.

Mathayas, N., Brown, D. E., Wallon, R. C., & Lindgren, R. (2019). Representational gesturing as an epistemic tool for the development of mechanistic explanatory models. *Science Education, 103*(4), 1047–1079.

Mezari, A., & Maglogiannis, I. (2017). Gesture recognition using symbolic aggregate approximation and dynamic time warping on motion data. In *the Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare* (pp. 342–347). Barcelona, Spain: ACM. doi:10.1145/3154862.3154927

Nathan, M. J., & Walkington, C. (2017). Grounded and embodied mathematical cognition: Promoting mathematical insight and proof using action and language. *Cognitive Research: Principles and Implications, 2*(1), 9. doi:10.1186/s41235-016-0040-5

Noice, H., & Noice, T. (2006). What studies of actors and acting can tell us about memory and cognitive functioning. *Current Directions in Psychological Science, 15*(1), 14–18.

Ochoa, X. (2017). Multimodal learning analytics. In *The Handbook of Learning Analytics* (pp. 129–141). doi:10.18608/hla17

Price, S., Jewitt, C., & Sakr, M. (2016). Embodied experiences of place: a study of history learning with mobile technologies. *Journal of Computer Assisted Learning, 32*(4), 345–359.

Prieto, L. P., Sharma, K., Kidzinski, Ł., Rodríguez-Triana, M. J., & Dillenbourg, P. (2018). Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data. *Journal of computer assisted learning, 34*(2), 193–203.

Segal, A., Tversky, B., & Black, J. (2014). Conceptually congruent actions can promote thought. *Journal of Applied Research in Memory and Cognition, 3*(3), 124–130.

Shapiro, L. (2019). *Embodied cognition*. New York, NY: Routledge.

Shen, S., & Chi, M. (2017, June). Clustering student sequential trajectories using dynamic time warping. In *the Proceedings of 10th the International Conference on Educational Data Mining (EDM)* (pp. 266–271), Wuhan, China: International Educational Data Mining Society.

Skulmowski, A., & Rey, G. D. (2018). Embodied learning: introducing a taxonomy based on bodily engagement and task integration. *Cognitive research: principles and implications, 3*(1), 6. doi:10.1186/s41235-018-0092-9

Smith, C., King, B., & Gonzalez, D. (2016). Using multimodal learning analytics to identify patterns of interactions in a body-based mathematics activity. *Journal of Interactive Learning Research, 27*(4), 355–379.

Stieff, M., Lira, M. E., & Scopelitis, S. A. (2016). Gesture supports spatial thinking in STEM. *Cognition and Instruction, 34*(2), 80–99.

Tretter, T. R., Jones, M. G., Andre, T., Negishi, A., & Minogue, J. (2006). Conceptual boundaries and distances: Students' and experts' concepts of the scale of scientific phenomena. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 43*(3), 282–319.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & review, 9*(4), 625–636.

Appendix A: Sample questions from pre- and post-tests

Conceptual Knowledge

- What causes earthquakes?

Exponential Knowledge

- What happens to the amount of damage in a town if the earthquake goes from a magnitude of 7.2 to a magnitude of 8.2?

Transfer Knowledge

- I want you to imagine that you McDonald's and Burger King are going to start opening restaurants in China. McDonald's plans to open 3,000 restaurants every year for 12 years. Burger King is going to start with 1 restaurant and then triple the number of restaurants every year for 12 years. Which restaurant chain do you think will have the most restaurants in 12 years?

Birt, J., & Vasilevski, N. (2021). Comparison of Single and Multiuser Immersive Mobile Virtual Reality Usability in Construction Education. *Educational Technology & Society*, *24* (2), 93–106.

Comparison of Single and Multiuser Immersive Mobile Virtual Reality Usability in Construction Education

James Birt[*] and Nikolche Vasilevski
Faculty of Society and Design, Bond University, Australia // jbirt@bond.edu.au // nvasilev@bond.edu.au
[*]Corresponding author

ABSTRACT: Immersive virtual reality (IVR) and mobile technologies have been identified as important in reimaging information delivery and pedagogy. This, coupled with evolving research in single (SUVR) and multiuser (MUVR) IVR environments, may enhance educational practice. However, there is limited research on the impact of such technologies on the learners' experience in authentic learning environments, such as building information modeling in architecture, engineering, and construction (AEC) workflows. This paper addresses this through a study of forty-eight participants recruited from a postgraduate construction course at an Australian University to answer a research question on how mobile MUVR is more useable than mobile SUVR when experiencing building information models. A within-subjects' experiment was performed using a mixed-methods approach assessing participant mobile IVR Usability on a 5-point Likert scale across four constructs and analysis of reflective sentiment and essays. The results show that when the participants moved from SUVR to MUVR, this significantly increased the overall perceived mobile IVR Usability. Combined with the qualitative analysis, these results suggest that MUVR influences mobile IVR Usability and an increase in learner experience. This study can be used as a launchpad for future research that will explore the causes of the evolution of the enhancement that MUVR provides, expanding beyond the scope of AEC education and industries.

Keywords: Virtual reality, Mobile, Multiuser, Learning experience, Virtual reality usability

1. Introduction

Building Information Modelling (BIM) is a new technique being used in the architecture, engineering, and construction (AEC) industry to assist with stakeholder communication, phase scheduling, and design error clash detection (Chan, Ma, Yi, Zhou, & Xiong, 2018). Used throughout the whole building lifecycle, BIM enables advanced forms of interactive visualization through object-related quantitative and qualitative data and visual 3D models (Santos, Costa, & Grilo, 2017). This combination of information and spatial modelling enables optimized workflows and communication that supports the critical success factors (Antwi-Afari, Li, Pärn, & Edwards, 2018) and ever-changing AEC industry compliance (Chan et al., 2018). With the data in AEC projects growing in scale and complexity (Tang, Shelden, Eastman, Pishdad-Bozorgi, & Gao, 2019) there has been a rapid shift in AEC education to integrate BIM into course content leading to integration difficulties (Puolitaival & Forsythe, 2016).

Immersive Virtual Reality (IVR) (Parong & Mayer, 2018; Radianti, Majchrzak, Fromm, & Wohlgenannt, 2020) and the subset of mobile IVR (Ladendorf, Schneider, & Xie, 2019; Olmos, Cavalcanti, Soler, Contero, & Alcañiz, 2018) has been identified as emerging technologies for reimagining education delivery. Coupled with evolving research in multiuser IVR environments (Buck, Rieser, Narasimham, & Bodenheimer, 2019), mobile IVR may enhance learning in spatial disciplines such as AEC education (Birt & Cowling, 2018; Vasilevski & Birt, 2020). However, there is limited empirical research on the usability impact of such methods on the learners' experience.

This paper addresses this through a study of forty-eight participants recruited from a postgraduate construction course at an Australian Higher Education institution. Using a mix of quantitative and qualitative data, participants answered the research question: "How is mobile MUVR more useable than mobile SUVR when experiencing building information models?"

2. Background

Unifying learner skills within professional real-world environments is difficult and complex (Herrington, Reeves, & Oliver, 2014). This is due to a combination of factors such as safety concerns, staffing costs, educating on non-existent environments, liability issues, repeatable experiences, and situated contextual feedback. With AEC projects growing in scale and complexity, there has been a rapid shift in AEC education to integrate authentic industry methods such as BIM into courses, leading to pedagogical integration difficulties.

Puolitaival and Forsythe (2016) noted that most AEC courses use traditional methods such as plan drawings to teach integrated BIM. However, urgent reform is required to embed emerging technologies with collaborative experientially driven approaches to improve authentic learning and industry readiness of learners.

Schott and Marshall (2018) recognize that we best understand experiential education as philosophy, which accepts the idea that learners need to interact with the world to comprehend it. Therefore, if cognitive immersive learning environments are achievable using technology which the visualization and simulation literature suggests (Bower, Lee, & Dalgarno, 2017; Dalgarno & Lee, 2010; Fowler, 2015; Reisoğlu, Topu, Yılmaz, Karakuş Yılmaz, & Göktaş, 2017), then the challenge is achieving alternatives to situated real-world learning through a combination of task design and interactive, immersive environments.

Increasingly IVR has been shown to assist in both industry (Slater & Sanchez-Vives, 2016) and educational (Jensen & Konradsen, 2018) settings. Radianti et al. (2020) published a systematic review of IVR and the pedagogical theories, methods, and domains in which IVR has been embedded in higher education. The authors mapped IVR technology into three domains: (i) mobile VR (Samsung Gear VR, Oculus Go/Quest), (ii) high-end heads mounted displays (HMD) (Oculus Rift, HTC VIVE) and, (iii) enhanced VR (often referred to as mixed reality using haptics and physical world data). It is noted that other systems such as 2D monitors, CAVEs, and similar are classified as non-immersive VR.

According to Radianti et al. (2020) the characteristics of IVR and learning are one of interactivity, presence, immersion, and the interplay of technical attributes such as graphical quality, framerate, and field of view on the individual sense of awareness associated to their personality traits. The authors noted that immersion is the most subjective of the characteristics. Slater and Sanchez-Vives (2016) view the discussion of immersion exclusively related to the IVR system's technical affordances and that immersion may give rise to subjectivity in users and is, therefore, a subjective correlation to presence. That is, if the participants perceive the use of their body within the IVR in a more natural way, then the body's perceptual system is perceived as being there. Johnson-Glenberg (2018) concludes that full-body movement and gestures provide for even more creative actions and learning in 3D with higher presence correlating with higher embodiment levels. This is further examined in their later work, where VR platform choice affects embodiment and learning (Johnson-Glenberg et al., 2020).

Dalgarno and Lee (2010) highlight the impact of immersion and fidelity on the participants' higher active learning function showing that as immersion increases, so does the learning. Fowler (2015) argues that immersion may emerge as a product of the complex interactions within the environment. Bailenson (2018) notes that the interactivity aspects of VR offer tremendous learning opportunities. The connection between immersion and engagement within IVR has been researched by a number of researchers (Allcoat & von Mühlenen, 2018; Hudson, Matson-Barkat, Pallamin, & Jegou, 2019). Huang, Backman, Backman, McGuire, and Moore (2019) noted that "virtual environments have the potential as an educational platform for real-world simulations, professional training, synchronous interaction, and global collaboration to provide interactive and immersive learning experiences and to enhance learner engagement."

Parong and Mayer (2018) identify IVR potential for engagement but cautioned on the immersive medium, as IVR may cause learners to be less reflective and that the VR may be better as a pre-lesson to spark interest, and then learning is formed around conventional collaboration. This is further examined by Bower et al. (2017) by exploring the importance of collaborative participant discussion (communication) and interaction. An essential question in understanding how to develop and build collaborative or multiuser IVR recognizes how people perform actions together, and the impact of social dynamics and affordance behavior is often studied through virtual avatars (Buck et al., 2019).

There is much literature on the use of virtual avatars for education (Dickey, 2005; Hew & Cheung, 2010; Mikropoulos & Natsis, 2011; Reisoğlu et al., 2017) and the impact of physical avatar characteristics on the participants deeper immersion and emotional satisfaction within IVR (Steed, Pan, Zisch, & Steptoe, 2016; Waltemate, Gall, Roth, Botsch, & Latoschik, 2018). The avatar's degree of personalization, realism, and graphical fidelity is significantly related to ownership, presence, and agency in immersive virtual environments (Waltemate et al., 2018). It has also been shown that avatar mediated communication has clear correlations to the social presence (Oh, Bailenson, & Welch, 2018). However, there are concerns about the impact that users' avatars have on others (Turkay & Kinzer, 2014) with deidentification (Merola & Peña, 2009) and motion control (Heidicker, Langbehn, & Steinicke, 2017) being essential considerations, especially in social, collaborative VR.

Heidicker et al. (2017) found that presence, social presence, and cognitive load are influenced by the avatars' body representation with full human representations increasing sense of presence. The degree of avatar mapping to users' movements increased the feeling of co-presence and behavioral interdependence. Waltemate et al. (2018) also found that the degree of HMD immersion had significant correlation effects in participant agency and feeling of presence.

The focus of this study is on mobile solutions. Current smartphone HMD technology that enables mobile IVR is limited in immersing a user completely in computer-mediated reality but does enable everyone to experience IVR (Ladendorf et al., 2019; Radianti et al., 2020). Papachristos, Vrellis, and Mikropoulos (2017) compared high-end HMD with mobile VR and concluded that mobile-based systems could provide acceptable immersion and contribute to the pedagogy. This is reiterated by Lai et al. (2019) and Olmos et al. (2018) with caution around graphical techniques, communication via audio, and a lack of gesture via reverse kinematics in mobile avatar visualization.

Birt and Cowling (2018) showed that emerging technologies using evolving experiential delivery using mobile devices and IVR show positive usability results in communicating design models across single user implementations. Vasilevski and Birt (2020) extended this research and explored a qualitative thematic analysis approach using assessment of learner reflections of mobile multiuser IVR. However, both stopped short of presenting quantitative assessment and correlation data of IVR usability and experience. This paper addresses this through a mix of quantitative and qualitative data.

3. Method

As noted in the literature review, unifying learner skills within professional real-world environments is difficult, especially when the authentic environments are immersive, complex, and collaborative. This study involves delivering a learning lesson focusing on building space and construction by integrating a BIM model delivered through single-user and multiuser mobile IVR.

3.1. Simulation design

A BIM model illustrated in Figure 1 showing site planning, floor plan, elevations, and sections was developed in Autodesk REVIT 2017.

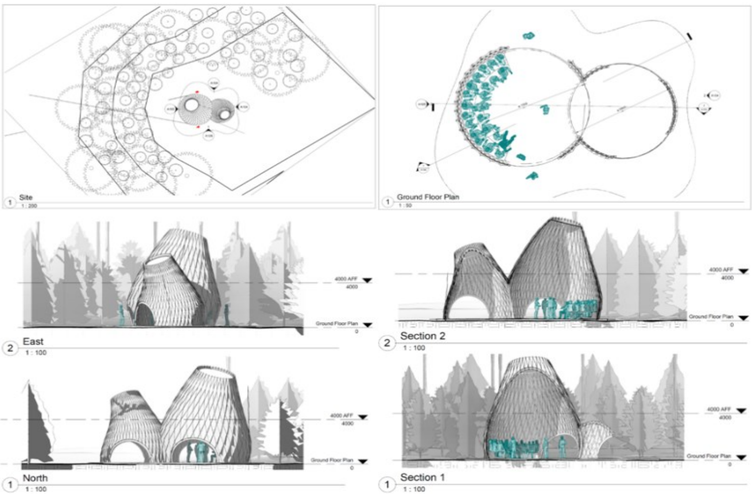


Figure 1. Model construction, site (top left), floor (top right), elevations (bottom left), and sections (bottom right)

The model was revised several times through consultation with the authors, course convener, and architectural designer. Using the developed BIM visualization, the lead author translated the model into mobile IVR using the game engine Unity and the Samsung Gear VR navigation plugin inbuilt into Unity. Model materials and lighting were constructed in line with the BIM model. However, due to mobile device performance constraints, realism was unfortunately reduced, as illustrated in Figure 2 with the BIM renders shown on the left and mobile IVR on the right. The visualization allows for observation of the BIM model over a simulated 24-hour transition simulating the day-night lighting change.

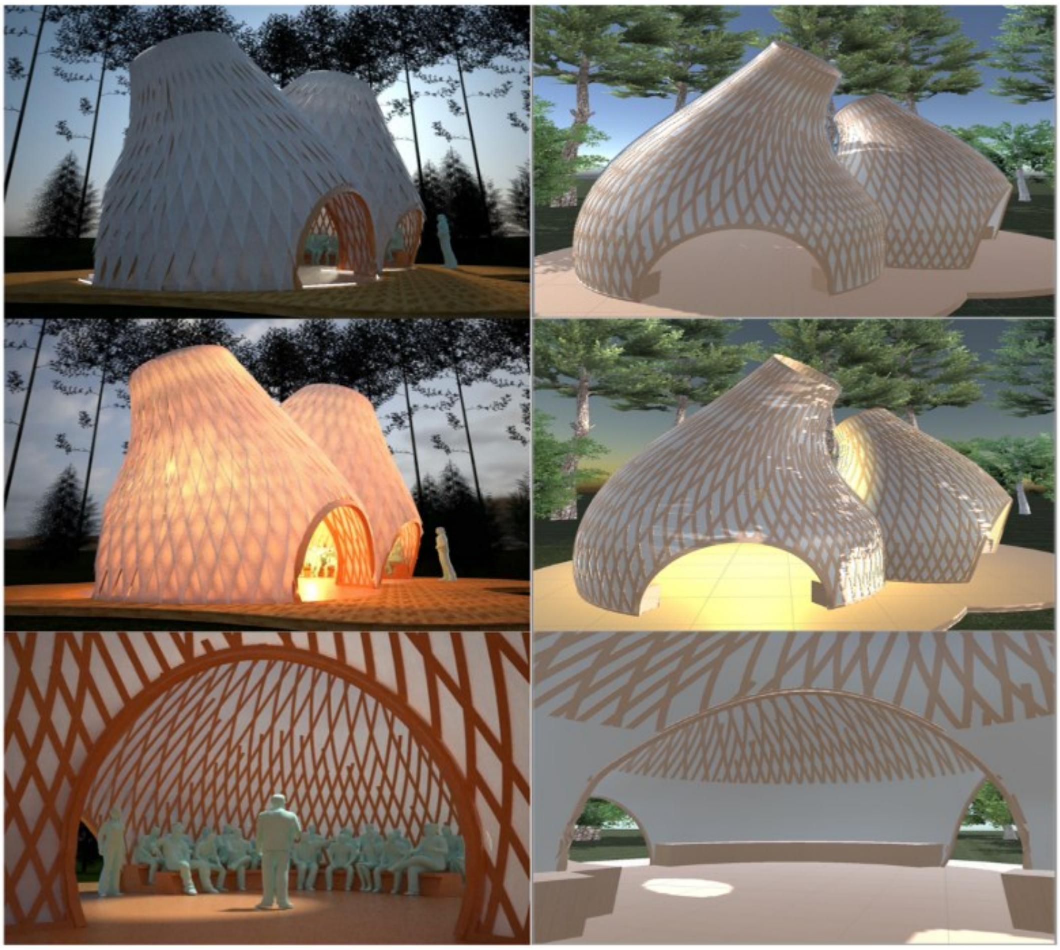


Figure 2. Rendered BIM (left), Unity IVR (right) highlighting mobile visualization compromises

The avatar was developed in two versions, see Figure 3, one with a head (used for the multiuser interaction), and one without (participant self-perspective).

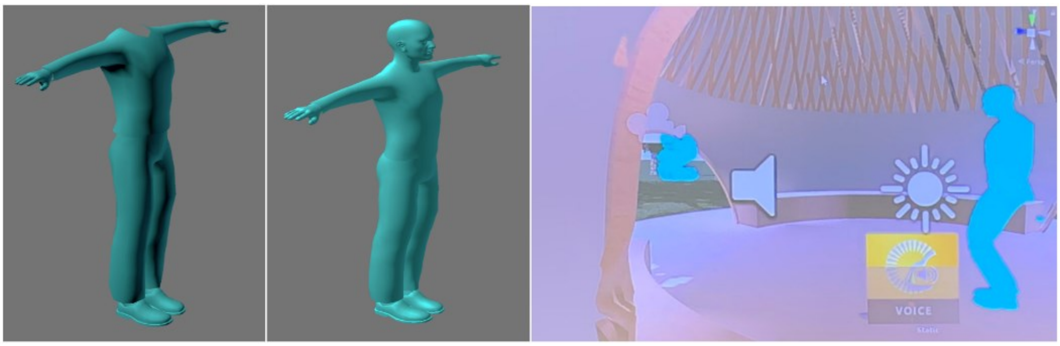


Figure 3. Constructed avatar showing self-view (without head), peer-view (with head), and visualization.

The avatar was built in line with Heidicker et al. (2017) that highlighted the importance of animation and user mapped motion control on the presence and behavioral interdependence. Although the authors understand the significance of avatar personalization, realism, and graphical fidelity (Steed et al., 2016; Waltemate et al., 2018), the overheads of development for a mobile IVR solution is too high. Therefore, a humanoid avatar was developed that identified the importance of complete body representations. The avatar was de-identified and colorized light blue to reduce participant impact effects (Merola & Peña, 2009; Turkay & Kinzer, 2014) but it is noted that the representation is of a male avatar, which represents a limitation of the research. The constructed MUVR environment included multiuser participant visualization of avatars and voice communication in line with IVR development (Bailenson, 2018; Dalgarno & Lee, 2010; Hew & Cheung, 2010; Reisoğlu et al., 2017). The device networking, multiple avatar handling, and voice chat were enabled through the Photon Networking version 1 (see https://www.photonengine.com/pun) plugin for Unity. The multiuser IVR environment was the same as the single-user environment to isolate only the variables of multiuser voice and multiuser avatar visualization of other participants to reduce effect bias (see Figure 4 for SUVR and Figure 5 for MUVR).

Figure 4. Single user experience with self-view (left -hand side), and classroom experience (right-hand side)

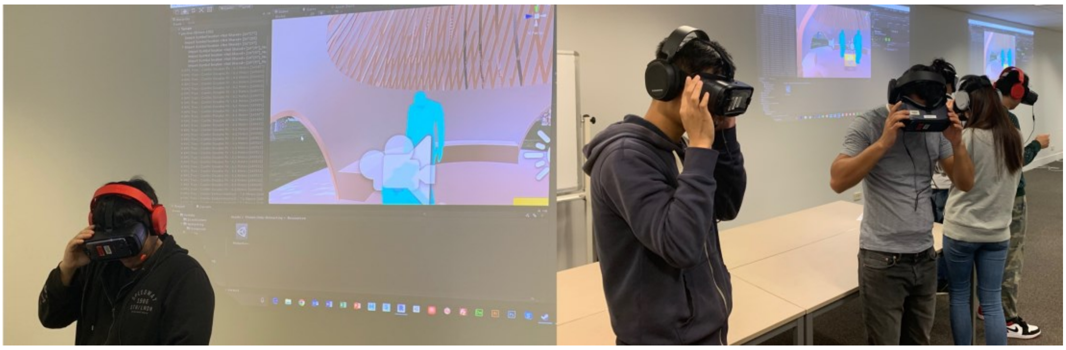


Figure 5. Multiuser experience with avatar kinematics (left-hand side), and classroom experience (right-hand side)

3.2. Study design

The study was performed as a within-subjects experiment using a mixed-methods research methodology per Creswell and Clark (2018) and the ethics granted for the study from the authors' institution. Data was collected from forty-eight (*n* = 48) voluntary recruited postgraduate construction participants from an Australian Higher Education institution in 2019. These participants were enrolled in a professional portfolio subject in the last semester of their Master of Construction degree. All data were recorded, stored, and de-identified under Australian Human Research Ethics.

Participants were randomly allocated into groups of 4-6 participants. After completing the single user condition, participants were moved to the multiuser condition. They were all provided with Samsung Galaxy S8 smartphones running Android Pie and Samsung Gear VR headsets with Steel Series 1 headsets. We selected these devices due to availability and to provide a consistent experience for the learners. Onboarding was presented for 5 minutes, delivered by the lead author. This included safety operation, assembly, task use, mobile IVR navigation, and communication method. The differences between the two experiences of the single user and multiuser were explained, including variance in the visual environment that being the use of simulated avatars and method of collaboration being the avatars and voice communication.

For both the single user (see Figure 4) and multiuser (see Figure 5) experiences, the participants were given 10 minutes per simulation to explore/communicate in the mobile IVR environment, 5 minutes per simulation to discuss their experience together as per Parong and Mayer (2018), and 20 minutes per simulation to complete an online Qualtrics survey (discussed in detail below). There was an additional 5-minute break between experiments to allow for transition and rest time.

3.3. Participant demographics

The participant demographics across the two experimental instances are presented in Table 1.

Table 1. Frequencies and percentages of participant demographics and technology competence

	Frequency	%		Frequency	%
Sex			Employment		
Male	20	42	Yes	2	4
Female	27	56	No	27	56
Undisclosed	1	2	No, but used to be	10	21
			Undisclosed	9	19
Age (mean = 26.7, median = 26.5)					
18-24	11	23	Technology Competence		
25-34	36	75	Extremely competent	6	13
35-44	1	2	Moderately competent	18	38
			Slightly competent	10	21
Country of residence			Neither competent nor incompetent	4	8
China	33	69	Slightly incompetent	1	2
India	5	10	Undisclosed	9	19
Australia	1	2			
Undisclosed	9	19	Primary mobile platform		
			Android (Google)	7	15
			iOS (Apple)	32	67
			Undisclosed	9	19

3.4. Data collection and instrument validation method

Self-reported measures in the form of an adapted questionnaire (see Table 2) were used to measure IVR usability and experience. The questions were adapted from the Nelson/Norman Group (see https://www.nngroup.com/articles/usability-testing-101/) leaders in software usability and Manis and Choi (2019) Virtual Reality Hardware Acceptance Model (VR-HAM), which is an extension to the original Technology Acceptance Model (TAM) proposed by Davis (1985).

Table 2. Immersive mobile virtual reality usability survey items

Construct	Item	Questionnaire item
IVR Utility	UTL1	The visualization is accessible or available for use at any time
	UTL2	The visualization is affordable in terms of monetary cost or efficiency in terms of time
	UTL3	The visualization is responsive, robust, and stable (error-free) in use (e.g., there are no problems with motion sickness, frame rates, or general software bugs)
IVR Engagement	ENG1	The novelty, aesthetics, or feedback afforded by the visualization focuses attention and involvement on the learning objective
	ENG2	The visualization is motivating, making the learner want to complete the learning objective
IVR Experience	EXP1	The visualization allows the accomplishment of the learning objective
	EXP2	The visualization allows for interactive variable manipulation, e.g., rotation, time, scene objects.
	EXP3	The visualization provides confidence in meeting the learning objective
	EXP4	The visualization provides effective or ease of re-establishing proficiency of the learning objective after a period (length) of time of the activity
	EXP5	The visualization allows for spatial translation (movement) of your (the users) viewpoint
	EXP6	The visualization provides a clear interface design to observe (view) and interpret the learning objective
	EXP7	The visualization provides an accurate representation of the real world (including visual, touch, and sound)
	EXP8	The visualization supports the discussion of learning objectives between stakeholders (instructor, learners, others)
	EXP9	The visualization allows emergent, creative, playful discovery towards the learning objective

All items were measured on a five-point Likert scale from strongly disagree to strongly agree. This addresses the concerns of subjective personality traits, as highlighted by Radianti et al. (2020) by quantifying the participants’ perception of mobile IVR Usability. The survey instrument was administered as an online Qualtrics questionnaire per ethics via QR code using a unique identifier.

The analysis of the validity and reliability of the measurement model and the analysis of the path model were undertaken using the component-based PLS-SEM (Smart-PLS 3.3.0) software. The model includes three reflective constructs as three dimensions of mobile IVR usability, given that indicators are assumed to be caused by the latent variable. These are mobile IVR Utility with three items, mobile IVR Engagement with two items, and mobile IVR Experience with nine items.

We assessed the validity and reliability of the measurement model by conducting a confirmatory factor analysis (CFA) to assess reliability, convergent validity, and discriminant validity of the constructs. We assessed convergent validity with three metrics: average variance extracted (AVE) and composite reliability (CR), and Cronbach's Alpha. The instrument's convergent reliability was evaluated through Cronbach's alpha (Alpha) at 0.95. The results exceeded the minimum value proposed by Nunnally (1978) with a Cronbach's alpha > 0.7.

The inter-correlation between constructs was not above 0.9, which is in line with Pavlou, Liang, and Xue (2007). The AVE of mobile IVR Utility (0.619), mobile IVR Engagement (0.678), and mobile IVR Experience (0.663) were higher than 0.5 (Fornell & Larcker, 1981). The composite reliability (CR) values were between 0.762 and 0.952, higher than 0.7 (Fornell & Larcker, 1981). See Table 3 for more details. With the above said, we conclude that the criteria for convergent validity are met.

Table 3. Convergent validity and reliability of the mobile IVR usability model

Variable	Cronbach's Alpha	rho_A	Composite Reliability	Average Variance Extracted (AVE)
IVR Engagement	0.750	0.781	0.762	0.619
IVR Experience	0.949	0.952	0.950	0.678
IVR Utility	0.853	0.862	0.854	0.663
IVR Usability	0.951	0.955	0.952	0.589

Regarding the discriminant validity, the inter-correlation of constructs did not exceed the square root of the AVE of either of those compared constructs see Table 4. The square root of the AVE of the three dimensions of mobile IVR Usability is 0.787, 0.823, and 0.814; thus, we can conclude that the discriminant validity is met. The mobile IVR Usability analysis was performed via one-way repeated measures ANOVA and two instances of one-sample *t*-test. The analysis of the individual constructs was conducted via correlation and frequencies comparison.

Table 4. Discriminant validity of the mobile IVR usability model

	IVR Engagement	IVR Experience	IVR Utility	IVR Usability
IVR Engagement	0.787			
IVR Experience	0.889	0.823		
IVR Utility	0.527	0.765	0.814	
IVR Usability	0.925	1.034	0.861	0.767

As per Creswell and Clark (2018) the quantitative data was also reflected upon with follow up qualitative data. This was achieved using both a participant opinion sentiment of the visualization methods collected through the Qualtrics survey and follow up reflective essays. The participant sentiment was collected at the time of the experiment which is presented in the results below. The 300-500 words reflective essay was submitted anonymously by the participants fourteen days after the experiment and linked to their course learning outcomes which is used in combination with the quantitative data as per the mixed methods approach.

4. Results

Data were analyzed in the IBM SPSS statistics package version 26.0.0.1. A comparison of IVR Usability in both single (SUVR) and multiuser (MUVR) is presented in Figure 6. The Usability means and adjusted 95% confidence intervals, as noted by Loftus and Masson (1994) for the SUVR, and MUVR conditions are displayed.

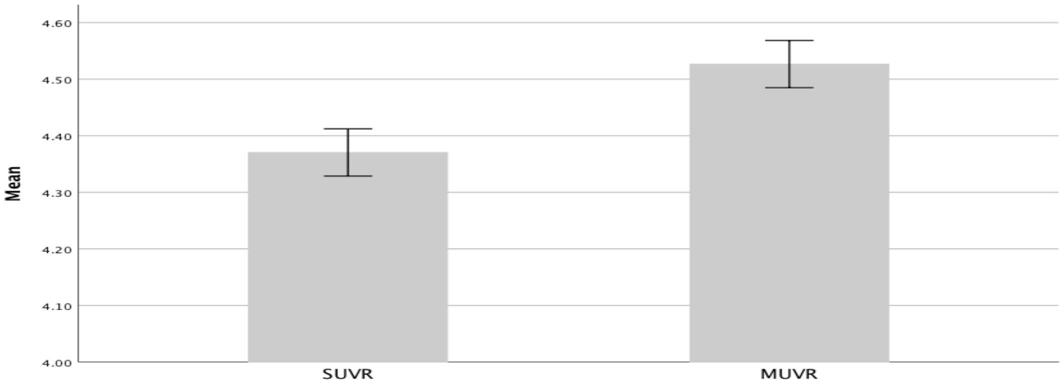


Figure 6. Mobile IVR Usability means and 95% confidence intervals for SUVR and MUVR

To test the hypothesis that for SUVR (M = 4.408, SD = 0.447) and MUVR (M = 4.564, SD = 0.434) mobile IVR Usability means were not equal, we used a one-way, repeated-measures ANOVA, followed by post-hoc one-sample t-tests to test the mobile IVR Usability difference to the baseline and the intervention measurements. Prior to conducting the analysis using parametric tests, we verified the normality of the distribution using a z-test for normality test using skewness and kurtosis (Kim, 2013). We obtained the z-score by dividing the skewness g = -0.522 (SE = 0.343) and kurtosis = -1.114 (SE = 0.674) values their standard errors resulting in values of and -1.278 and -.165, respectively. Both values were under the threshold of 1.96 (at $\alpha = 0.05$). We consider that the assumption of normality of the sample is satisfied.

The assumption of homogeneity of variance was tested (Morgan, 1939; Pitman, 1939) based on a t-test method found in Gardner (2001) and satisfied at $t = 0.747$ (at $p < .05$). The assumption of normality for the t-test was considered satisfied as the skewness and kurtosis were at 0.513 and 2.715, respectively, which is less than the thresholds allowed for t-test (skewness <|2.0| and kurtosis <|9.0|) as per (Posten, 1984).

A repeated-measures ANOVA determined a significant effect of MUVR on the mobile IVR Usability of the visualization, $F(1, 47) = 14.18, p < .001, \eta^2 = .232$. These results suggested that the visualization was significantly more usable in terms of mobile IVR when used in a multiuser environment. Regarding the individual dependent variables, a repeated-measures ANOVA for mobile IVR Utility showed significant main effects for the MUVR intervention, $F(1, 47) = 11.34, p = .002, \eta^2 = .194$. This is also the case with the mobile IVR Experiences construct, which also showed significant main effects for the MUVR intervention, $F(1, 47) = 11.33, p = .002, \eta^2 = .194$. However, the mobile IVR Engagement construct did not show a significant main effect for the MUVR intervention.

A one-sample one-tailed t-test was run to determine whether the SUVR score in recruited participants was greater than neutral (neither agree nor disagree), defined as an SUVR score of 3 on the Likert scale. Mean SUVR score (4.393 ± 0.444) was higher than the normal SUVR score of 3.0, a statistically significant difference of 1.393 (95% CI, 1.267 to 1.519), $t(47) = 22.20$, $p < .001$. Thus, the SUVR mean was statistically significantly higher than the threshold of 3.

A one-sample one-tailed t-test was also run to determine whether the MUVR score in recruited subjects was greater than neutral (neither agree nor disagree), defined as a MUVR score of 3. Mean MUVR score (4.504 ± 0.518) was higher than the neutral MUVR score of 3, a statistically significant difference of 1.504 (95% CI, 1.357 to 1.651), $t(47) = 20.55$, $p < .001$. Therefore, the MUVR mean was statistically significantly higher than the measured mobile IVR usability threshold of 3. The SUVR and MUVR survey results are presented in Table 5, showing the comparison of the construct variable means and the correlations between the pairs. A paired-samples t-test of all the items was performed to compare items across all the constructs and presented in Table 6.

Table 5. Comparison of constructs means and correlations between SUVR and MUVR pairs

Pair	SUVR			MUVR			Correlations		
	Mean	*SD*	Err	Mean	*SD*	Err	MDif	Cor	Sig.
IVR Engagement	4.464	0.486	0.075	4.560	0.471	0.073	-0.095	0.355	.021
IVR Experience	4.362	0.418	0.064	4.553	0.441	0.068	-0.191	0.648	< .001
IVR Utility	4.111	0.578	0.089	4.310	0.692	0.107	-0.198	0.840	< .001
IVR Usability	4.323	0.412	0.064	4.502	0.429	0.066	-0.179	0.744	< .001

Table 6. Paired sample *t*-test of construct items highlighting differences between the two conditions per item

Pair	Mean	*SD*	Std. Error Mean	*t*	*df*	Sig. (2-tailed)
UTL1	0.063	0.727	0.105	0.596	47	.554
UTL2	-0.146	0.505	0.073	-2.001	47	.051
UTL3	-0.437	0.741	0.107	-4.09	47	0
EXP1	-0.167	0.663	0.096	-1.741	47	.088
EXP2	-0.083	0.577	0.083	-1	47	.322
EXP3	-0.062	0.633	0.091	-0.684	47	.497
EXP4	-0.208	0.617	0.089	-2.338	47	.024
EXP5	-0.104	0.722	0.104	-1	47	.322
EXP6	-0.167	0.519	0.075	-2.224	47	.031
EXP7	-0.458	0.922	0.133	-3.446	47	.001
EXP8	-0.125	0.57	0.082	-1.52	47	.135
EXP9	-0.125	0.444	0.064	-1.952	47	.057
ENG1	-0.104	0.592	0.085	-1.219	47	.229
ENG2	-0.062	0.633	0.091	-0.684	47	.497

Correlations were tested between the MUVR constructs against age and technology competence, and the significant correlations are presented in Table 7.

Table 7. Significant Correlations between MUVR mobile IVR Usability constructs against age and technology competence (Spearman's correlation coefficient is denoted by r_s)

Construct	r_s	Sig. (2-tailed)	*N*
What is your age group?			
IVR Engagement	0.390	0.014	39
IVR Experience	0.420	0.008	39
IVR Usability	0.400	0.012	39
How competent do you consider yourself when using technology?			
IVR Engagement	-0.324	0.044	39

The results of the collected qualitative SUVR and MUVR reflective opinion (sentiment) comments were collected at the time of the experiment. The constructs of mobile IVR Usability are presented as frequencies of participants that commented on the concept (or synonyms) in Table 8. The similarities between the visualization methods in the comments were evident, including a common sentiment of IVR Engagement, IVR Utility and IVR Usability across both methods. The primary difference was in IVR Experience for the MUVR approach. Qualitative data from the reflective essays is presented below in the findings and discussions to inform the quantitative data as per the mixed methods approach.

Table 8. Frequencies of the participant reflective opinions of SUVR and MUVR mapped to constructs

Pair	SUVR		MUVR		
	Count	%	Count	%	
IVR Engagement	11	23%	8	17%	engaging, interesting, fun, cool, exploring, want to
IVR Experience	8	17%	15	31%	experience, feel, communication, voice, people
IVR Utility	12	25%	9	19%	use, practical, useful, good to use, easy to control, helpful
IVR Usability	24	50%	22	46%	good, great, nice, fantastic, awesome, perfect, amazing, impressive, better
Negative	7	15%	8	17%	dizzy, price, slow, or any other negative comment

5. Findings and discussions

To answer the research question, we measured mobile IVR Usability of both the SUVR and MUVR visualizations, comparing results using parametric statistical tests addressing the subjectivity concerns of Radianti et al. (2020). It is noted that the comparison between SUVR and MUVR is a repeated measure (within-subjects) comparison as per the experimental design. In both visualizations, participants performed well and similarly regarding the mobile IVR Usability. The one-sample *t*-tests showed that both the SUVR and MUVR visualization methods are usable in terms of mobile IVR Usability. This is further supported by the participant sentiment data shown in Table 8.

The ANOVA of the repeated measures showed that the MUVR results are statistically significantly higher than the SUVR, although not explicitly shown in the qualitative sentiment data collected at the time of the experiment. Subsequently, the strong positive correlation usability between SUVR and MUVR suggests that if the multiuser mode is enabled in the IVR system, the mobile IVR Usability of the system is expected to increase. This is consistent with the research on education use of IVR and the impact of communication, multiple avatar representation, and learners' experiences (Bailenson, 2018; Jensen & Konradsen, 2018; Radianti et al., 2020; Reisoğlu et al., 2017).

We conducted a comparison between the constructs' items and the independent variables, and the significant correlations are shown in Table 7. The results for age were in positive correlation to several construct items. As noted in previous literature by Manis and Choi (2019), there is typically a negative relationship between increasing age and perceived ease of use and usefulness as it relates to VR hardware. We argue that one of the reasons might be the very narrow age range for the majority of the participants (8 years), with 98% of the sample being between 23 and 31years old, and 76% between 23 and 28. To further unpack *how* mobile MUVR compares to SUVR in terms of IVR Usability, we discuss the constructs below.

5.1. Mobile immersive virtual reality engagement

The mobile IVR Engagement construct did not perform significantly better in MUVR than SUVR, which we argue is the result of varying technology competency between the randomly assigned participants. The participants that self-report as more competent with technology tend to be less engaged within the MUVR environment, as the correlation results suggest. Therefore, this can result in a conflict between the participants' motivation to complete the learning task as those that self-report as more competent are grouped with those that self-report as less competent.

However, the qualitative essay data suggests that MUVR is more engaging and motivating. Specifically, students share their reflections of the excitement they had, "This is definitely another exciting day." the sensory appeal "…VR technologies can alter our senses of the reality. Interaction played a major part under this construct, as found in statements like "it is amazing to feel that we are interactive to the model and other group members." The observation during the activity also supports that students felt more engaged during the MUVR activity. Students found the visualization to be motivating and made them want to learn and explore more, which is supported by the statements like, "It's really fun, that makes me have more interests to learn." This aligns with research in authentic learning that authentic environments maintain high levels of engagement (Herrington et al., 2014; Parong & Mayer, 2018) and that immersive and interaction leads to engagement (Allcoat & von Mühlenen, 2018; Huang et al., 2019; Hudson et al., 2019).

5.2. Mobile immersive virtual reality experience

The mobile IVR Experience in MUVR was significantly better for the participants when compared to SUVR. This is supported by both the sentiment and qualitative reflection data from the survey and essays.

First, regarding the accomplishment of the learning objective, which was spatial understanding of BIM models, the observation during the hands-on activities and the essays submitted show that the students found using the VR visualizations helped them learn. This is reflected in student comments from the MUVR experience, such as "…makes me have more interest to learn" and "Very active learning method, a good way to demonstrate your idea." We also explored the satisfaction with the visualization. This is illustrated by statements such as "we experienced the fantastic [VR] used in the construction industry" and "I gained a great deal of information about the BIM course and had a good experience using VR." The satisfaction of the MUVR experience is captured in the student comments: "the overall effects and video chatting were parts of this fascinating experience." This reflects the broader literature that IVR improves learning and leads to greater satisfaction (Jensen & Konradsen, 2018; Reisoğlu et al., 2017).

When examining the deeper learning or proficiency of remembering and retaining the knowledge after the activity, students reflected on the experience. This demonstrates memory retention and remembering intricacies of the visualizations in detail, even though the reflections were written two weeks after the activity. Parong and Mayer (2018) suggest that using VR may result in losing the capacity to retain memory and reflect. In line with that, we argue that the relatively short continuous VR times and having the learner to reflect after each activity in the survey and comment may have led to better retention evident in the essays.

Interaction is a core IVR characteristic (Radianti et al., 2020; Slater & Sanchez-Vives, 2016). Interaction is not unique to the MUVR instance and is found in many of the students' statements, such as "It really immersed me into that virtual world and saw clearly how it rendered and how it looks in different time of the day." However, of note was that interaction with other students was even more important to the students in MUVR, supported with statements like "…it still impresses me that I can interact with peers and study together …" and "it is amazing to feel that we are interactive to the model and other group members." This is in line with Dalgarno and Lee (2010), who suggest that immersive virtual environments lead to an increase in co-presence or user's perception of being there.

The movement of the user's viewpoint in the visualization is also linked to the interaction. The freedom of movement in any virtual world is essential. It relates to the perception of control, immersion, and engagement (Jensen & Konradsen, 2018; Slater & Sanchez-Vives, 2016). Even though the visualization offered only three degrees of freedom, students found the ability to move via their animated avatars in the IVR environment very engaging and fun, by using statements such as "In the scene, we can walk around at will…". The multiuser additions through avatars and voice added to this experience, supported with comments like, "…we can feel each other in the model, and it's funny that all of us jumped on to the tree outside the [pavilion]." This reflects the consensus that avatars lead to improved immersion (Oh et al., 2018; Steed et al., 2016).

Concerning the visualization providing a clear interface design for observation and interpretation, overall the visualization performed satisfactorily, which is supported by comments such as "In terms of multiuser [the visualization] will give a clear understanding collectively" and "It really immersed me into that virtual world …" All essay comments related to this item were found only in MUVR suggesting that co-presence was a factor in the participants' observation and interpretation. We note that both visualization environments were virtually the same. However, as the MUVR visualization was enhanced through interactive avatars and communication between the participants, this likely resulted in an increased perceived fidelity, as the inclusion of the multiuser features brings it closer to the real world. Participants' comments support this in MUVR with statements like: "more real in details" and "Good shadow of leaf and tree." In SUVR, the statements are less optimistic towards the fidelity, for instance: "…can see the virtual world as real[.] However it's still not good enough…" and "Image quality should be improved." The lack of personalization of the avatars could be a negative contributing factor in line with Waltemate et al. (2018). In addition to this, participants commented: "…the body shape images are not beautiful." However, communication in multiuser may also have had a counter effect to the avatar beauty, by increasing the reality of the environment, supported by comments like: "Interestingly, we can "walk" in this VR environment by moving our feet in reality, and when the audio function turned on, I could hear someone talking in the [pavilion]."

In the reflective essays, students relate communication to almost all other concepts, using comments such as, "It was observed that while doing it individually it was less effective but when it was done in group it created a better understanding of collaborative use of this technology." Students found the peer to peer communication in a multiuser VR environment to be very important in construction, by stating: "[VR technologies] … they make people communicate better." and "Effective communication is the essential part for any construction project." It was strongly expected that this feature would be dominant in the MUVR, which is supported by the fact that almost none of the students talk about it in a single-user context. This suggests that the conversational aspects of multiuser IVR environments, as discussed by Bower et al. (2017) and the social presence (Oh et al., 2018) represented by avatars may have an effect on learner communication and their experience within the IVR environment.

Concerning the emergent, creative, and/or playful discovery allowed by the visualization, students, in general, found the VR environment to be fun and engaging: "It's very cool to stand on roof.", "It's really fun ..." and "Makes me feel like exploring more." Overall, the visualization supported students' creativity. Some students perceived MUVR more fun and some even as game-like, stating: "The feeling in multiplayer model is like playing the multi players games or online games thought the VR." Perceiving the MUVR experience in a gameful way can increase the engagement and further motivate the learners in meeting the learning objective (Dalgarno & Lee, 2010).

5.3. Mobile immersive virtual reality utility

In exploring the utilitarian dimension of mobile IVR, MUVR performed significantly better when compared to SUVR. Even though the same interface was used, the addition of the avatars, voice communication, and social interaction may have led to perceiving the system as more complex, and because there were no slowdowns or crashes, better to use. The issues with motion sickness that some of the participants experienced were not unique to either of the instances. Concerning the responsiveness, the arguably slow mobile devices were not observably different between the instances. The addition of the social element may have mitigated or lessened some of the adverse effects, evident in comments such as “when [used] for more than 5min I feel a little bit headache. But it is really f[u]n to see everyone [on] the same place”. This would be in line with the research of Oh et al. (2018) and the role of avatars and social interactions on presence.

6. Study limitations

There are limitations within the study, with only subjective self-reported measures used, a population sample of 48 learners within a single subject, three degrees of freedom head-mounted displays in the form of the Samsung Gear VR, now a discontinued technology, and a narrow age band. There are limitations with the instrument design, as many questions have been merged to reduce the impact on the participant’s time to complete the questionnaire and should be expanded in future research. Counterbalancing should be done in future studies to reduce the carryover effect, including using a 2X2 design using different environmental conditions, but this would require more participants. Therefore, this study can be considered a launchpad for future research exploring the causes of the evolution of the enhancement that multiuser mobile IVR provides, expanding beyond the scope of AEC education. Future research should focus on larger population sample sizes, counterbalancing of groups, new six degrees of freedom mobile HMDs, and improved visual capacity of new hardware devices across different disciplines.

7. Conclusions

With BIM adoption increasing in the AEC industry, this has led to integration difficulty in educational practice. IVR and mobile pedagogy has been shown to assist in education and authentic learning environments. In this study, participants observed a BIM visualization over a 24-hour transition simulating the day-night lighting change using mobile IVR technology. Participants experienced both single and multiuser visualizations with the difference between the experiences being the addition of visual avatars and voice communication capability. The students were able to visualize not only their self-avatar but that of their peers in the virtual environment. This setup also facilitated enhanced affordances such as agency, perception, and peer learning between the students allowing for collaborative navigation and synchronous communication discussion of the environment in the multiuser vs. an asynchronous approach in the single-user experiment. This study found that the perceived mobile IVR usability of MUVR to be higher than SUVR. Implementing the multiuser functions in the experiment engaged the learners on a different level and enhanced the overall learning experience. Although the study revealed that multiuser facilitates a better learning environment for learners, we do not assert that using multiuser mobile IVR is always the best way given increased costs associated with development and deployment. Therefore, we propose integrating mobile IVR and specifically multiuser within the course offerings when appropriate and achievable.

References

Allcoat, D., & von Mühlenen, A. (2018). Learning in virtual reality: Effects on performance, emotion and engagement. *Research in Learning Technology, 26*. doi:10.25304/rlt.v26.2140

Antwi-Afari, M. F., Li, H., Pärn, E. A., & Edwards, D. J. (2018). Critical success factors for implementing building information modelling (BIM): A Longitudinal review. *Automation in Construction, 91*, 100-110. doi:10.1016/j.autcon.2018.03.010

Bailenson, J. (2018). *Experience on demand: What virtual reality is, how it works, and what it ca*n do. New York, NY: WW Norton & Company.

Birt, J., & Cowling, M. (2018). Assessing mobile mixed reality affordances as a comparative visualization pedagogy for design communication. *Research in Learning Technology, 26*. doi:10.25304/rlt.v26.2128

Bower, M., Lee, M. J. W., & Dalgarno, B. (2017). Collaborative learning across physical and virtual worlds: Factors supporting and constraining learners in a blended reality environment. *British Journal of Educational Technology, 48*(2), 407-430. doi:10.1111/bjet.12435

Buck, L. E., Rieser, J. J., Narasimham, G., & Bodenheimer, B. (2019). Interpersonal Affordances and Social Dynamics in Collaborative Immersive Virtual Environments: Passing Together Through Apertures. *IEEE transactions on visualization and computer graphics, 25*(5), 2123-2133. doi:10.1109/TVCG.2019.2899232

Chan, A. P. C., Ma, X., Yi, W., Zhou, X., & Xiong, F. (2018). Critical review of studies on building information modeling (BIM) in project management. *Frontiers of Engineering Management, 5*(3), 394-406. doi:10.15302/j-fem-2018203

Creswell, J. W., & Clark, V. L. P. (2018). *Designing and conducting mixed methods research* (3rd ed.). Thousand Oaks, CA: Sage publications

Dalgarno, B., & Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology, 41*(1), 10-32. doi:10.1111/j.1467-8535.2009.01038.x

Davis, F. D. (1985). *A Technology acceptance model for empirically testing new end-user information systems: Theory and results.* Cambridge, MA: Massachusetts Institute of Technology,

Dickey, M. D. (2005). Brave new (interactive) worlds: A Review of the design affordances and constraints of two 3D virtual worlds as interactive learning environments. *Interactive Learning Environments, 13*(1-2), 121-137. doi:10.1080/10494820500173714

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

Fowler, C. (2015). Virtual reality and learning: Where is the pedagogy? *British Journal of Educational Technology, 46*(2), 412-422. doi:10.1111/bjet.12135

Gardner, R. C. (2001). *Psychological statistics using SPSS for Windows*. Upper Saddle River, NJ: Prentice Hall.

Heidicker, P., Langbehn, E., & Steinicke, F. (2017). Influence of avatar appearance on presence in social VR. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, 233-234. doi:10.1109/3DUI.2017.7893357

Herrington, J., Reeves, T. C., & Oliver, R. (2014). Authentic Learning Environments. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (pp. 401-412). New York, NY: Springer New York.

Hew, K. F., & Cheung, W. S. (2010). Use of three-dimensional (3-D) immersive virtual worlds in K-12 and higher education settings: A review of the research. *British Journal of Educational Technology, 41*(1), 33-55. doi:10.1111/j.1467-8535.2008.00900.x

Huang, Y.-C., Backman, S. J., Backman, K. F., McGuire, F. A., & Moore, D. (2019). An Investigation of motivation and experience in virtual learning environments: A Self-determination theory. *Education and Information Technologies, 24*(1), 591-611. doi:10.1007/s10639-018-9784-5

Hudson, S., Matson-Barkat, S., Pallamin, N., & Jegou, G. (2019). With or without you? Interaction and immersion in a virtual reality experience. *Journal of Business Research, 100*, 459-468. doi:10.1016/j.jbusres.2018.10.062

Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies, 23*(4), 1515-1529. doi:10.1007/s10639-017-9676-0

Johnson-Glenberg, M. C. (2018). Immersive VR and education: Embodied design principles that include gesture and hand controls. *Frontiers in Robotics and AI, 5*(81). doi:10.3389/frobt.2018.00081

Johnson-Glenberg, M. C., Ly, V., Su, M., Zavala, R. N., Bartolomeo, H., & Kalina, E. (2020, June). *Embodied Agentic STEM Education: Effects of 3D VR Compared to 2D PC.* Paper presented at the 2020 6th International Conference of the Immersive Learning Research Network (iLRN), San Luis Obispo, CA.

Kim, H.-Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative dentistry & endodontics, 38*(1), 52-54.

Ladendorf, K., Schneider, D., & Xie, Y. (2019). Mobile-based virtual reality: Why and how does it support learning. In *Handbook of Mobile Teaching and Learning* (pp. 1-19). New York, NY: Springer.

Lai, Z., Hu, Y. C., Cui, Y., Sun, L., Dai, N., & Lee, H. (2019). Furion: Engineering high-quality immersive virtual reality on today's mobile devices. *IEEE Transactions on Mobile Computing*, 19(7), 1586 – 1602. doi:10.1109/TMC.2019.2913364

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic bulletin & review, 1*(4), 476-490. doi:10.3758/BF03210951

Manis, K. T., & Choi, D. (2019). The Virtual reality hardware acceptance model (VR-HAM): Extending and individuating the technology acceptance model (TAM) for virtual reality hardware. *Journal of Business Research, 100*, 503-513. doi:10.1016/j.jbusres.2018.10.021

Merola, N., & Peña, J. (2009). The Effects of avatar appearance in virtual worlds. *Journal for Virtual Worlds Research, 2*(5). doi:10.4101/jvwr.v2i5.843

Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A Ten-year review of empirical research (1999–2009). *Computers & Education, 56*(3), 769-780. doi:10.1016/j.compedu.2010.10.020

Morgan, W. (1939). A Test for the significance of the difference between the two variances in a sample from a normal bivariate population. *Biometrika, 31*(1/2), 13-19.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

Oh, C. S., Bailenson, J. N., & Welch, G. F. (2018). A Systematic review of social presence: definition, antecedents, and implications. *Frontiers in Robotics and AI, 5*(114). doi:10.3389/frobt.2018.00114

Olmos, E., Cavalcanti, J. F., Soler, J.-L., Contero, M., & Alcañiz, M. (2018). Mobile virtual reality: A Promising technology to change the way we learn and teach. In *Mobile and ubiquitous learning* (pp. 95-106). doi:10.1007/978-981-10-6144-8_6

Papachristos, N. M., Vrellis, I., & Mikropoulos, T. A. (2017). A Comparison between Oculus Rift and a low-cost smartphone VR headset: Immersive user experience and learning. In *Proceedings of the 17th International Conference on Advanced Learning Technologies (ICALT)* (pp. 477-481). doi:10.1109/ICALT.2017.145

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology, 110*(6), 785-797. doi:10.1037/edu0000241

Pavlou, P. A., Liang, H., & Xue, Y. (2007). Understanding and mitigating uncertainty in online exchange relationships: A Principal-agent perspective. *MIS Quarterly*, *31*(1), 105-136. doi:10.2307/25148783

Pitman, E. (1939). A Note on normal correlation. *Biometrika, 31*(1/2), 9-12.

Posten, H. O. (1984). Robustness of the two-sample t-test. In D. Rasch & M. L. Tiku (Eds.), *Robustness of Statistical Methods and Nonparametric Statistics* (pp. 92-99). Dordrecht, The Netherlands: Springer Netherlands.

Puolitaival, T., & Forsythe, P. (2016). Practical challenges of BIM education. *International journal of building pathology and adaptation, 34*(4/5), 351–366. doi:10.1108/SS-12-2015-0053

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. doi:10.1016/j.compedu.2019.103778

Reisoğlu, I., Topu, B., Yılmaz, R., Karakuş Yılmaz, T., & Göktaş, Y. (2017). 3D virtual learning environments in education: A Meta-review. *Asia Pacific Education Review, 18*(1), 81-100. doi:10.1007/s12564-016-9467-0

Santos, R., Costa, A. A., & Grilo, A. (2017). Bibliometric analysis and review of building information modelling literature published between 2005 and 2015. *Automation in Construction, 80*, 118-136. doi:10.1016/j.autcon.2017.03.005

Schott, C., & Marshall, S. (2018). Virtual reality and situated experiential education: A Conceptualization and exploratory trial. *Journal of Computer Assisted Learning, 34*(6), 843-852. doi:10.1111/jcal.12293

Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI, 3*(74). doi:10.3389/frobt.2016.00074

Steed, A., Pan, Y., Zisch, F., & Steptoe, W. (2016). The Impact of a self-avatar on cognitive load in immersive virtual reality. In *Proceedings of the 2016 IEEE Virtual Reality (VR)* (pp. 67-76). doi:10.1109/VR.2016.7504689

Tang, S., Shelden, D. R., Eastman, C. M., Pishdad-Bozorgi, P., & Gao, X. (2019). A Review of building information modeling (BIM) and the internet of things (IoT) devices integration: Present status and future trends. *Automation in Construction, 101*, 127-139. doi:10.1016/j.autcon.2019.01.020

Turkay, S., & Kinzer, C. K. (2014). The Effects of avatar-based customization on player identification. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS), 6*(1), 1-25. doi:10.4018/ijgcms.2014010101

Vasilevski, N., & Birt, J. (2020). Analysing construction student experiences of mobile mixed reality enhanced learning in virtual and augmented reality environments. *Research in Learning Technology, 28*. doi:10.25304/rlt.v28.2329

Waltemate, T., Gall, D., Roth, D., Botsch, M., & Latoschik, M. E. (2018). The Impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE transactions on visualization and computer graphics, 24*(4), 1643-1652. doi:10.1109/TVCG.2018.2794629

Holly, M., Pirker, J., Resch, S., Brettschuh, S., & Gütl, C. (2021). Designing VR Experiences – Expectations for Teaching and Learning in VR. *Educational Technology & Society*, *24* (2), 107–119.

Designing VR Experiences – Expectations for Teaching and Learning in VR

Michael Holly*, Johanna Pirker, Sebastian Resch, Sandra Brettschuh and Christian Gütl

Graz University of Technology, Austria // michael.holly@tugraz.at // johanna.pirker@tugraz.at // sebastian.resch@tugraz.at // sandra.brettschuh@tugraz.at // c.guetl@tugraz.at

*Corresponding author

ABSTRACT: Skills in science, technology, engineering, and mathematics (STEM) are increasingly in demand. Theoretical knowledge and formulas alone are frequently not sufficient to understand complex phenomena. Simulations are a valuable tool to support the conceptual understanding by visualizing invisible processes. The constant interaction with the learning material is an essential factor when learning with simulations and virtual worlds. Virtual reality (VR) technologies enable interaction with the virtual environment with a high intensity of immersion. Maroon is a VR platform for teaching physics and has been in development for over five years. Previous results with Maroon have already demonstrated the potential of virtual reality for learners and teachers, but also highlighted a list of potential challenges in terms of VR experience design, usability, and pedagogical concepts. Over the past six months, we have conducted user studies with a total of 85 participants, both student teachers (n = 26) and pupils (n = 59) at high schools and teacher training institutions. In this paper, we want to facilitate the difficult task of designing educational VR platforms by describing the expectations of educators and pupils.

Keywords: STEM education, Virtual Reality, Interactive simulations, Immersive learning

1. Introduction

Innovative technologies and high-quality scientific research have a significant impact on our daily lives. Topics in the field of science, technology, engineering, and mathematics (STEM) are becoming more and more relevant and are a significant driver of innovation. Consequently, there is a growing demand for experts with experience, know-how, and skills in these fields (Zeidler, 2016). Olson and Riordan (2012) have already pointed out the need to increase the number of students with a degree in STEM disciplines. Lack of interest and enthusiasm are among the reasons for high failures in these fields. Students describe it as boring, complicated and uninteresting. It is not clear to many students why they must study natural sciences. It is therefore necessary to promote the interest and motivation of students in the STEM disciplines (Reeve, Jang, Carrell, Jeon, & Barch, 2004).

As Freedman (1996) shows, teaching science is a challenging task, especially in the field of physics. Traditional teaching methods present solutions and concepts, but they fail to teach how to solve problems. Hake (1998) confirms this observation by showing that students have difficulties understanding conceptual aspects while memorizing formulas. An old Chinese saying supports the concepts of learning through experience: “I hear, and I forget. I see and I remember. I do and I understand.” The integration of educational activities which involves learners in the learning process has shown to be a successful teaching method. Students learn by doing things and think about what they are doing (Freeman et al., 2014). Learning tools with interactive and engaging activities can help students to better understand such conceptual aspects. Sanders (2008) shows that in physics, the combination of hands-on experiments, interactive simulations and active participation are valuable tools that support learners during their educational process. Interactive simulations such as PhET (see https://phet.colorado.edu/) allow students to take the ownership of the learning experience and support their conceptual understanding by making connections to everyday life. They can be integrated into teacher demos, interactive discussions, classroom activities, labs and homework to support teachers in illustrating concepts (Moore, Chamberlain, Parson, & Perkins, 2014). This active learning concept can be extended by various modern technologies to meet the needs of a new generation of learners in a flexible and digital way. Simulations, visualizations, virtual and remote laboratories support students in self-directed, active, and group-based learning.

De Jong et al. (2013) showed that any lab form has its advantages for certain use cases. While real laboratories are more suitable for acquiring hands-on experience, virtual laboratories enable expandable experiments, multiple access and visual representation of unseen phenomena with minimal potential for the occurrence of dangerous situations. In virtual laboratories designed for active learning , learners become part of the simulated environment by interacting with the virtual world, which helps them to learn complex concepts. The combination of simulations and visualizations in a lab-like environment offers schools and universities a cost-effective way to provide learning experiences similar to those in real-world labs (Asıksoy & Islek, 2017). The use of immersive and interactive technologies such as virtual reality opens new possibilities for creating engaging learning experiences.

In previous works (Pirker et al., 2017a; Pirker et al., 2017b; Pirker, Lesjak, Parger, & Gütl, 2018) we introduced Maroon, an interactive physics laboratory and experiment environment designed to teach physics in a more engaging and immersive way. In small-scale studies we found that various VR technologies can be used to support and engage learners in understanding physics. Maroon supports immersive VR, where a stereoscopic head-mounted display (HMD) that is tracked by sensors is used. With the HMD and controllers it is possible to achieve spatial immersion and to conduct hands-on experiments. Additionally, Maroon can be used on a PC (desktop VR), where a three-dimensional world is simulated on-screen. In a recent study (Pirker et al., 2019a) we collected qualitative data from secondary school teachers to identify use cases, design goals and issues. In this paper we intend to extend our previous work (Pirker, Holly, & Gütl, 2020) to facilitate the difficult task of designing educational VR platforms by describing the challenges and the expectations of educators and pupils. The main research goals are defined as the following:

- Identification of educator and learner expectations for teaching and learning in VR.
- Providing a first guideline for the design of a VR learning environment for the classroom.
- Discussion of challenges and recommendations for the design and development of learning and teaching activities in VR.

Contribution: In this paper, we present a study with 26 student teachers and 59 pupils, discussing the room-scale VR version of the learning environment Maroon for physics education in classroom situations. The focus is on identifying and discussing expectations for teaching and learning in VR and challenges by combining the study results with a literature review.

The following section gives an overview of the background and the related work in the field of STEM education and of the challenges in designing educational VR platforms. In Section 3, we introduce the virtual physics laboratory Maroon and the conceptual design of the experiments and simulations. Section 4 describes the study design, and Section 5 presents the results. In Section 6, we discuss the challenges and recommendations based on the literature review and our findings. Section 7 closes with a discussion about implications, potential and gives some ideas for further studies.

2. Related work

STEM subjects have high drop-out and failure rates. This can be related to difficulties students have in understanding theoretical concepts (Olson & Riordan, 2012). Therefore, researchers such as Olson and Riordan (2012) or Dori, Hult, Breslow, and Belcher (2007) combined traditional classroom experiences with interactive and engaging digital learning experiences such as virtual simulations or animations to help students better understand the underlying concepts and phenomena. Conducting experiments in virtual or remote laboratories that would otherwise be dangerous, hard to conduct and/or too expensive is another way for effectively support learners (Corter et al., 2007). Wieman and Perkins (2005) showed that digital resources are a cost-effective, safe, and fast alternative to traditional learning methods and experiment setups. Furthermore, these digital resources have been shown to help users to achieve a better understanding.

VR tools can support the students in their learning process. Slavova and Mu (2018) showed that when using VR as a complementary tool to traditional learning methods, students showed higher performance in understanding and recognizing concepts. Furthermore, learners using VR technologies remain more motivated during their learning process (Liou & Chang, 2018). Bogusevschi, Muntean, and Muntean (2020) have published a study on the effect of a virtual 3D physics learning environment on 12-13-year old students that concluded that over 74% of the participants found the simulation helpful for gaining a better understanding.

Due to the decreasing prices of HMDs, we could observe that an increasing number of studies on the use of VR, especially for educational use, have been published. Furthermore, HMDs have become lighter and even wireless alternatives, such as the Oculus Quest, are now available, factors which are helping to increase their broad distribution and use. Due to the lower costs of HMDs, the increased research and the technical process, VR is becoming both more immersive and more widely available to a wider user base (Dempsey, 2016), making it easier to use VR learning experiences as an addition to the traditional classroom setups.

Nevertheless, researchers are still investigating the challenges that arise from the use of VR as a learning tool. Velev and Zlateva (2017) point out that a highly immersive experience must be created to achieve a high learning rate. According to Stark (1995) VR does not depend on highly realistic surroundings, as long as the virtual space offers enough cues for the perceptual system, immersion can be achieved. It is of greater importance when creating an immersive experience that the motion sickness is reduced to a minimum and the interface design of the VR simulation is appealing (Callaghan, Eguíluz, McLaughlin, & McShane, 2015). Especially in the field of physics, immersive learning experiences help to achieve a better understanding, particularly if the design places a special emphasis on interactivity (Pirker, Gütl, Belcher, & Bailey, 2013; Pirker, Berger, Guetl, Belcher, & Bailey, 2012). Abulrub, Attridge, and Williams (2011) confirm that for a high learning rate, the VR experience must be designed so that the user can actively interact with the virtual world.

Furthermore, not only the setup time of HMDs and virtual simulations for the use in real classrooms could become problematic, but also the time needed for students to become familiar with the new technology (Velev & Zlateva, 2017). They need a guided tutorial or someone to help them and guide them through their first experiences (Abdelaziz, Alaa El Din, & Senousy, 2014; Safikhani, Holly, & Pirker, 2020). De Jong and Van Joolingen (1998) describe that it has a positive influence on the learning effect when theoretical information is available during the simulation. Additionally, they explain that tasks help the users to focus on the desired outcome and guide them through the experience. Most teachers have limited time and classroom settings usually provide limited hardware resources, factors which reduce the time each of their students has in the virtual laboratory. Most VR learning experiences allow only limited personal interactions, which makes collaborative learning and teamwork difficult (Velev & Zlateva, 2017).

One of the most challenging, but also most promising parts of VR learning experiences is the optimization of the knowledge acquisition process. Mainly Callaghan, Eguíluz, McLaughlin, and McShane (2015), Velev and Zlateva (2017) and Liou and Chang (2018) discovered various guidelines for designing VR learning experiences. Creating a good virtual experience requires that users are not initially overwhelmend with the virutal world, which can be achieved by reducing dizziness and motion sickness. Furthermore, VR laboratories offer many possibilities that conventional ones cannot provide. For example, invisible phenomena can be visualized and made interactable and students can perform dangerous and expensive experiments. Students are offered a fully controllable and safe learning environment that reduces health risks, in case of performing dangerous experiments, and ensures that the same input will always lead to the same output. It also gives learners the possibility to apply their theoretical knowledge in practice, which motivates them and boosts their creativity. Given the large number of studies focusing on VR learning environments and laboratories, Ip and Li (2015) criticize that most of them failed to demonstrate an increase in long-term knowledge acquisition and that more long-term observations should be made.

3. Maroon – Virtual learning application

Maroon (see https://maroon.tugraz.at; to put someone ashore and abandon them on an island) is an interactive virtual physics laboratory that allows students to explore various experiments and phenomena in an immersive and engaging way. It is implemented in Unity (see https://unity.com) and supports different platforms with different levels of immersion such as virtual reality, mobile devices, or web-based applications. The learning activities and experiments are designed for active learning to involve students in the learning process. The main laboratory room (Figure 1) consists of different experiment stations and functions as a three-dimensional menu where the user can select one of the experiments by navigating to the specific station. Currently, the laboratory contains a whiteboard scene with different learning lessons, and eight experiments in the field of electromagnetism, electrostatics, oscillation, and waves (Pirker et al., 2019b). The experiments support several virtual learning experiences with different forms of engagement and immersion through diverse activities and interactions. In the following subsections the application concepts of Maroon VR and the different experiment setups are presented.

Figure 1. Overview of Maroon's laboratory with different learning stations

3.1. Maroon VR

The VR version of Maroon extends the core functionalities of Maroon with room-scale VR support for the HTC Vive (see https://www.vive.com) and the Oculus Rift (see https://www.oculus.com/rift). This version of Maroon offers a virtual reality experience with a high degree of immersion and focus on the learning content. Users can walk freely in the classroom within a predefined play area, allowing natural movement within the virtual environment. Due to real-world restrictions and the limitation of the tracking area, users need to become familiar with a different form of movement to cover greater distances in the virtual world. In VR applications, teleportation has become a standard and allows the user to move to a certain position while minimizing the feeling of motion sickness. For teleportation, the user can press a button on the controller to activate a colored arc, by moving the controller the user can point to the desired destination. After releasing the button, the user is placed in that position. Teleport markers in front of each experiment station help to navigate to a specific experiment by capturing the teleport beam. Each experiment or activity can be started from an entry-point that acts as a portal into the simulation room. Each experiment has a customized user interface and offers various virtual control elements to control specific experiment parameters and visualizations.

3.2. Experiment setups in VR

Maroon VR supports two different types of experiment setups each with an individual design. The first experiment setup contains only the elements necessary for the simulation (see Figure 2a). The experiment is placed in the middle of a simple and compact room and can be controlled via virtual control panels on the right and left side. This allows the user to observe the experiment results while setting the experiment parameters. To keep the interaction as simple as possible, all elements are arranged in such a way that they are accessible without use of teleporting. In contrast, the second setup contains additional gamified elements in a retro-futuristic laboratory room with a simplified design to improve student's engagement and motivation (see Figure 2b). The predefined quest list gives clear instructions by displaying different tasks, the progress, and additional information about the simulation.

(a) (b)

Figure 2. Different design setups in VR: a) simple experiment room with virtual control elements and b) retro-futuristic laboratory room with quest management system

3.3. Experiments and simulations

In this section the two interactive experiments Faraday's Law and Huygens's Principle are presented, which were used to evaluate the room-scale VR version of Maroon. The goal of these simulations is to simulate and visualize the concept of induction and diffraction. Each of them implements the general physical phenomenon and extends it with an appropriate user interface. The Faraday's Law Experiment shows the principle of induction when interacting with a permanent magnet and a conductive non-magnetic ring. Whenever the user moves the magnet, it causes a change in the magnetic flux and induces an electric current that generates another magnetic field. Through different visualizations the invisible magnetic field becomes visible and helps the user to better understand the underlying concepts. The virtual controls allow the user to influence the result of the experiment by changing the parameters of the magnet and the coil. The special feature of this experiment is the haptic feedback from the controllers, which allows the user to feel the acting forces. The controller vibrates as soon as a physical force acts on the magnet, where low vibration means a weak force, and high vibration means a heavy force. This allows users to have a real feeling of the acting forces. The Huygens's Principle Experiment uses water waves in a basin to demonstrate the physical concept of diffraction. It is a phenomenon that occurs when a wave hits an obstacle or a slit. To show the effect of diffraction, a slit plate is placed into the basin. When a wave hits this plate, the points on the wave act as a new source of secondary waves that propagate. This results in an interference pattern behind the plate. To obtain different interference patterns, the user can replace the plate with three types of slit plates. The experiment is influenced by the user by grabbing and moving the plates and changing physical parameters such as frequency, amplitude, wavelength, or the propagation mode. To make the wave peaks and wave trough more visible, the wave color can be changed using a color wheel.

4. Evaluation

In previous studies, we focused on different learning experiences with room-scale VR, mobile VR, and traditional screen-based technologies as well as on engagement, usability, and user experience with room-scale VR as the most engaging and most immersive form. However, we also found that teachers and students have very different opinions and expectations about experience design in VR. Therefore, in this paper, we focus in this paper on identifying teacher and pupil expectations for teaching and learning in VR. Since there are known issues convincing experienced teachers to try new things in educational technology, we conducted a user study with 26 student teachers and 59 pupils, which are open for new technologies. The study was organized in cooperation with two local schools and two universities with a focus on the following:

- Experience and engagement,
- Usability,
- Learning value from learner's perspective,
- Learning value from teacher's perspective.

4.1. Setup

We used two portable setups, including a gaming notebook, an HTC Vive HMD, two controllers, two lighthouses, and two tripods for the lighthouses. The two setups were placed in a single classroom with a minimum size of 2m x 2m for each VR station. All participants were in the same classroom and worked in pairs. Half of the participants tested the simple design of the experimental room, and the other half tested the retro-futuristic laboratory room. While the persons who tested the experiments in VR were given instructions from their partners outside (Figure 3a), the others tried the same experiments in the desktop version (Figure 3b). After 20 minutes, the groups were swapped, so that the participants using the VR version were now using the PC version and vice versa. To get a real classroom situation the setup was conducted within the context of a physics class and in a university classroom where the student teachers were role-playing as pupils. During the tests participants were not wearing earphones so we could talk to them and guide them through the experiments.


(a)


(b)

Figure 3. Classroom setup: a) VR workstation and b) desktop workstation

4.2. Material and procedure

For this study, we worked with two high schools and two universities to test Maroon from the perspective of students and prospective teachers. At the beginning of each test run the participants were asked to fill out a pre-questionnaire to obtain information about their ages, gender, experience with games, VR platforms, and e-learning tools. They were then given a short introduction how to work and interact with the environment. We introduced them to the controls and gave them instructions for the different tasks to be completed during the test run.

Pupils and student teachers were asked to perform the following tasks:

- Introductory session: Have a look at the lab environment and get a first impression.
- Go to the Faradays' Law experiment and start the simulation by moving the magnet. Try to identify the relationship between the electrical current and the acting force. Find out how the parameters of the magnet and the coil affect the experiment outcome.
- Go to the Huygens' Principle experiment and start the simulation. Try to understand the concept of diffraction and describe the interference pattern behind the slit plate.
- Take time to try other experiments (optional).

We provided an additional exercise sheet for the pupils in which they could work on different questions about the experiments. After conducting the experiments, the participants were asked to fill out a post-questionnaire in which they had to answer open-ended questions about their overall experience; 22 questions on a Likert scale between 1 (fully disagree) and 5 (fully agree) about their sentiment towards the physics lab, and 10 questions regarding usability. We used the System Usability Scale (SUS) (Brooke, 1996) to measure the system usability and the Computer Emotion Scale (Kay & Loverock, 2008) to assess users' experiences of interacting and learning with the virtual environment. To gain a deeper understanding of the user experience and use cases, pupils and student teachers were interviewed about the experiments and application scenarios in school classes. Student teachers were also asked about pedagogical models and cooperative scenarios.

4.3. Participants

In total 40 high school pupils, 19 pupils from an engineering secondary school (51 male, 8 female) and 26 student teachers (15 male, 11 female) took part in the user study. All 26 student teachers were attending the teacher education program for physics. The pupils were aged 13 to 19 (AVG = 14.41, SD = 1.87) and the student teachers from 21 to 31 (AVG = 23.69, SD = 2.65). 14 pupils and 11 student teachers had a visual impairment (including people wearing glasses).

We asked each of them to rate their experience with computers, video games and VR on a Likert scale from 1 (low) to 5 (high). Most pupils rated themselves as experienced with computers (AVG = 3.22, SD = 0.91) and video games (AVG = 3.59, SD = 1.16). Student teachers stated that they were also experienced in using computers (AVG = 3.12, SD = 0.91) but rated their experience with video games lower (AVG = 2.54, SD = 1.65). Pupils and student teachers indicated that they often play video games (pupils: AVG = 2.68, SD = 1.41; teachers: AVG = 2.27, SD = 1.46). 52 of the pupils and 16 of the student teachers liked playing video games but had little experience with VR (pupils: *AVG* = 1.69, *SD* = 0.93; teachers: *AVG* = 1.50, *SD* = 0.86). 52 pupils and 19 student teachers had heard about VR devices previously, but only 22 pupils and 7 student teachers had tried one. 45 pupils and 20 student teachers have already used an e-learning tool. Both pupils and student teachers consider the use of virtual reality in physics lessons to be a good idea.

5. Results

In this section, we describe the results obtained from the open-ended answers and the answers based on a Likert scale. Since the learning outcome depends on the user experience and the acceptance of the system, we focus on usability, engagement, immersion, and the learning experience as well as on application scenarios and pedagogical models.

5.1. Usability and user experience

Most participants were able to handle the VR controllers without any problems. Only a few of them had some initial problems when they tried to move in the virtual world or wanted to interact. They teleported themselves against walls or used the wrong button for interaction. By contrast, all participants were able to interact with the PC version without additional instructions. The users described the desktop application as more familiar but would prefer the interaction in VR as it is more realistic and natural. We used the Computer Emotion Scale to evaluate the emotions anger, anxiety, happiness, and sadness when learning with the VR environment. As shown in Table 1, pupils and student teachers rated the emotion happiness (e.g., satisfied, excited, curious) as high and the emotions of sadness, anger, and anxiety as very low. In total 44 pupils and 23 student teachers have completed the SUS questionnaire. It shows the degree of usability from 0 (poor) to 100 (excellent) and was rated with 76.5 by pupils and with 73.5 by student teachers. The resulting scores indicate an above average usability compared to other VR learning tools. When we asked participants whether they felt nauseous or dizzy while using VR, only two pupils and one student teacher reported cyber sickness.

Table 1. Results of the 12-item Computer Emotion Scale on a Likert Scale between 0 (never) and 3 (always)

	Pupils		Student Teachers	
	AVG	*SD*	*AVG*	*SD*
Satisfied	2.63	0.48	2.58	0.69
Excited	2.78	0.41	2.73	0.44
Curious	2.73	0.48	2.73	0.44
Happy	2.46	0.56	2.42	0.74
Depressed	0.15	0.48	0.04	0.19
Discouraged	0.36	0.73	0.04	0.19
Scared	0.22	0.49	0.08	0.27
Insecure	0.54	0.62	0.69	0.72
Helpless	0.41	0.49	0.46	0.57
Nervous	0.51	0.74	0.38	0.56
Frustrated	0.03	0.18	0.08	0.27
Angry	0.05	0.39	0.04	0.19

5.2. Immersion and engagement

Pupils rated the immersion of the experience on a Likert scale from 1 (not immersive at all) to 10 (fully immersive) as an average of 9.00 (*SD* = 0.98). Some pupils mentioned that learning with Maroon was easier and more engaging than traditional learning methods. Only two pupils experienced dizziness, which could hinder a convincing user experience.

Prospective student teachers rated the immersion of the experience on the Likert scale from 1 to 10 slightly lower than the pupils with an average of 8.50 (*SD* = 1.24). Among other things, it was noted that Maroon VR has a high degree of interaction that allows the inclusion of multiple senses. Furthermore, some of the student teachers mentioned the feeling of forgetting about being in a virtual world after spending some time within Maroon, which was induced by the high degree of interaction.

5.3. Learning value from the learner's perspective

To evaluate the learning experience, pupils were asked to rate their learning experience on a Likert scale between 1 (not agree) and 5 (fully agree). Table 2 gives an overview of the pupils' results with a focus on learning experiences. The pupils indicated that the experience was more engaging (*AVG* = 4.17, *SD* = 0.91) and fun (*AVG* = 4.68, *SD* = 0.63). In general, learners found that the lab made the content more interesting and easier to understand. Most of them said they would like to learn with Maroon (*AVG* = 4.39, *SD* = 0.89).

They also reported that they liked the immersive laboratory and outlined that concepts are easier to understand using three-dimensional visualizations and interactions than with traditional methods. Being able to see field-lines that are usually not visible contributed to this. They also mentioned that compared to real life experiments it was easier to conduct the experiments in VR and that Maroon was a fun experience and was a welcome relaxation from their typical school routine. Suggestions for improvements made the experiments look even more realistic, further improving the visualizations. A larger collection of experiments was also mentioned to improve Maroon.

5.4. Learning value from the teacher's perspective

Besides pupils, student teachers were also asked to rate the learning experience with Maroon on a Likert scale between 1 (not agree) and 5 (fully agree). Most student teachers mentioned that the VR setup is a good supplement to regular learning (*AVG* = 4.15, *SD* = 1.08). They also reported that Maroon makes the learning content more interesting (*AVG* = 3.88, *SD* = 1.18) and easier to understand (*AVG* = 3.65, *SD* = 1.13). In general, student teachers found that learning with Maroon was more motivating than ordinary exercises and more fun, as can also be seen from Table 2.

The student teachers reported that VR can be a good way to extend the variation of teaching methods in the classroom. The fact that Maroon appeals to multiple senses makes the system attractive for several types of learners, while the novelty of the technology is an attractive way to motivate students who normally show little interest in a subject. It was also mentioned that virtual experiments are valuable when the real experiment is expensive or dangerous, since students can carry these out with no health risks. Also mentioned was the possibility of visualizing unseen phenomena, learning in a playful way and the ability to change the experiments very quickly. Some student teachers criticized Maroon as being too close to regular games in its current state, and this makes it more difficult to keep students focused. To allow students to use it without close monitoring and guidance, and to prevent students from being distracted, the system should include detailed tasks with clear instructions. It was also found that obtaining enough VR headsets for all students is currently too costly and for this reason would be best to use the technology in the form of projects days. The setup time for the system can be regarded as reasonable in the context of such a project day.

In addition, student teachers recommended the use of videos in Maroon to provide additional support in the form of introductory film material that revisits the learning topics discussed, or videos showing the experiment and additional information to support the process of learning while experimenting in VR.

Table 2. Learning Experience rated by pupils and student teachers on a Likert Scale between 1 (not agree) and 5 (fully agree)

	Pupils		Student Teachers	
	AVG	*SD*	*AVG*	*SD*
I would like to learn with Maroon.	4.39	0.89	3.54	1.27
It is a good idea to use Maroon for learning.	4.51	0.86	3.92	0.98
Maroon is a good supplement to regular learning.	4.56	0.79	4.15	1.08
I learned something with Maroon.	4.08	0.92	2.69	1.23
Maroon makes the content more interesting.	4.63	0.64	3.88	1.18
Maroon makes the content easier to understand.	4.27	0.89	3.65	1.13
Maroon makes learning more engaging.	4.17	0.91	3.19	0.85
Maroon makes learning more fun.	4.68	0.63	4.08	1.09
Maroon makes learning more interesting.	4.46	0.73	3.96	1.08
The experience with Maroon inspired me to learn more about physics.	3.49	1.22	3.12	1.21
Learning with Maroon was more motivating than ordinary exercises.	4.51	0.70	4.23	0.95
It makes course content more interesting to learn about.	4.44	0.77	3.62	1.10
I would rather like to learn Physics with Maroon than with traditional	4.14	0.96	2.81	1.27

methods.				
I find regular physics classes boring.	3.19	1.25	1.54	0.58
Seeing the simulations with the VR glasses was engaging.	4.17	0.81	3.81	0.98
Seeing the simulations with the VR glasses was interesting.	4.42	0.65	4.54	0.58
Seeing the simulations with the VR glasses was more engaging than without.	4.37	0.87	3.96	1.04
I would rather use Maroon on my phone (+ VR glasses).	3.34	1.36	2.85	1.22
I would rather use Maroon on my own PC.	3.81	1.07	3.12	1.24
I would like to learn with Maroon in the classroom.	4.41	0.87	3.77	0.99
I would buy the VR glasses and download Maroon at home.	3.81	1.12	3.04	1.46
It was interesting to use Maroon.	4.81	0.43	4.73	0.53

5.5. Use cases and pedagogical models

The perspectives of pupils and teachers are often quite different. To consider their ideas and suggestions on how to use VR in schools and what can be learned in VR into consideration, we asked the pupils and student teachers in open questions to describe their ideas and scenarios. Based on these answers, we provide a list of use cases and subjects that can be applied in schools.

5.5.1. How to use VR in schools

To identify different use cases, we asked pupils (P) and prospective student teachers (T) how they would use VR in school. Student teachers were also asked to give collaborative scenarios as well as pedagogical models.

The following use cases were mentioned:

- *Dedicated VR room (P, T)* – Pupils mentioned that they would prefer a dedicated classroom for learning with VR. This includes learning as an integral part of classroom lessons as well as after-school (extracurricular) activities to repeat and review the learned materials. The student teachers also suggested a dedicated room with pre-installed VR setups to save time.
- *Mobile VR experience (P, T)* – Due to the financial constraints, pupils and student teachers mentioned the usage of smartphones in combination with a mobile VR headset. That allows pupils to run experiments on their devices at school and at home. But are limited in their use due to the processing power of these devices.
- *Weekly classes (P, T)* – Some pupils suggested using VR on a weekly basis to supplement the learning material with simulations and visualizations. Student teachers also mentioned the potential of regular VR sessions in classes to increase the learning outcomes of pupils.
- *Optional Subject (P)* – Pupils also recommended an optional subject or an afternoon class where they can use VR for learning and increasing their skills.
- *Project/Group Work (T)* – Several student teachers reported the potential of blocked VR classes in the form of project days during which pupils can work in groups on different experiments.
- *Autonomous Learning (T)* – The student teachers mentioned that VR as an effective educational method, could be a valuable tool to support students in autonomous learning. They also mentioned that using worksheets might be useful to give pupils clear instructions and help them to focus on the experiment.
- *Collaborative Learning (T)* - Since pupils benefit from each other's resources and skills when learning together, student teachers suggested working in groups, where one pupil performs the experiment and the others give hints and take notes.

5.5.2. Teaching subjects

To identify additional subjects and experiments, we asked pupils and student teachers about phenomena they would like to see, learn, or teach in VR.

The following subjects were mentioned:

- Astronomy
- Physics
- Chemistry
- Biology

- History
- Engineering

While pupils were interested in different STEM subjects, the student teachers were mainly interested in experiments in their own field – physics. All of them also highlighted that VR would be an opportunity for performing expensive and dangerous experiments within class and for visualizing and experimenting with unseen phenomena.

6. Discussion

In this study, we examined the challenges and recommendations for learning and teaching in VR, with an exemplary focus on the physics domain. In previous studies (Pirker et al., 2017a; Pirker, Lesjak, Parger, & Gütl, 2018), we have evaluated different VR experiences versus a classic desktop experience with a small group of students and teachers. We showed the potential of VR in the field of STEM and discussed the advantages and disadvantages. In this paper, we aim to facilitate the challenging task of designing VR learning platforms by describing the challenges and giving some recommendations to overcome these challenges.

6.1. Challenges and recommendations

Summarizing the findings of our research and study, we concluded that we can distinguish between the following main challenge categories: (1) immersion, (2) costs, (3), time restrictions, (4) knowledge gaining process.

To achieve an increased learning effect, it is crucial to offer a highly immersive experience, since then the user becomes part of the virtual world. Moreover, motion sickness should be restricted to a minimum. We recommend using HMDs combined with motion tracking and input devices that enable the interaction with the virtual world, as this already ensures a high level of immersion. Furthermore, the user interface should be tailored to the VR world, meaning that we create an appealing interface design and give the learners the possibility to interact directly with their learning content. As a result, users show more motivation for the learning content and simulations as we enable them to learn and explore the phenomena in a playful way (Velev & Zlateva, 2017; Callaghan, Eguíluz, McLaughlin, & McShane, 2015). We observed that students could ask questions regarding controls and experiment specifics, because the headsets were not equipped with earphones. This was especially helpful for first-time VR users, because they could show where a problem was in VR and we could see what they were trying to accomplish on-screen and point them in the right direction.

Another factor for using VR experiences complementary to the traditional classroom learning methods are the included costs. VR equipment prices already decreased in the previous years and HMDs are now available to a broader user base. Nevertheless, it is still a significant factor for most schools as they cannot afford to buy many HMDs and equip whole classes (Abulrub, Attridge, & Williams, 2011). We thus recommend using VR setups within project days, so that less VR equipment will be needed. Using VR laboratories instead of traditional ones reduces the costs of laboratory equipment and gives learners the possibility to perform even expensive experiments within the virtual world as many times as they want. Such an approach presents itself as a possible solution until mobile VR headsets can provide a similar degree of immersion and computing capabilities.

When using VR, an introduction of some kind is generally required to familiarize the user with the HMDs and controls. We must consider the training time for users (Velev & Zlateva, 2017). This can be optimized by using guided tutorials to explain how to use the controls (Abdelaziz, Alaa El Din, & Senousy, 2014), not to mention that with tutorials teachers do not need to guide each student personally and hence more time is saved. In addition, the setup time factor is significant if the VR equipment is to be used during traditional lessons. Hence, we recommend using the virtual laboratories within project days or coordinate with other classes, so that the equipment is in use throughout the whole day. In consequence the VR laboratory offers the possibility to change experiment settings or reset the simulation with a few commands. In addition, there is no need to prepare experiments in the real world, this is especially useful when having very time-consuming experiments.

To create a successful VR learning experience the knowledge gaining process should be optimized. Crucial for the success is that users should not be overwhelmed when first entering the virtual world (Callaghan, Eguíluz, McLaughlin, & McShane, 2015). Furthermore, it is important that the students know what the goal of the simulation is and that they do not feel lost in the simulation. During the tests, we noticed that already having background knowledge is a significant advantage for performing simulations. As Mayer's cognitive theory of multimedia states, learners receive information via two separate channels (visual and auditory) with limited capacity. Learners must select and organize the relevant information and integrate it based upon prior knowledge (Mayer, 2002). We found that a predefined quest list with clear task instructions in combination with worksheets can help students to stay focused and motivated. Moreover, with the use of VR we encourage the motivation of learners in a playful way while keeping them safe during the performance of dangerous experiments (Callaghan, Eguíluz, McLaughlin, & McShane, 2015). Through the use of VR we can provide users with the possibility for visualizing invisible phenomena, which proved to make the learning content more readily comprehensible for students (Slavova & Mu, 2018). The greatest advantage, however, is that learners can explore and interact with the simulations at their own pace and this motivates them, encourages their creativity and lets them utilize their theoretical knowledge.

6.2. Limitations

Due to the classroom setup in schools, the number of participants and the time constraints, an A/B split user study was not feasible. We therefore decided to set up the study as a workshop where participants tested the VR environment in a classroom situation. The different room layouts, light conditions and the shape of play areas had a marginal influence on the tracking accuracy, which caused slightly different experiences. While some pupils had experience in programming, VR, and video games, others had not been much involved in computer games, VR, and programming during their curriculum. The learning effects were determined by self-evaluation and do not indicate long-term effects. Since not all participants (15 pupils and 3 student teachers) completed the questionnaire, some data sets could not be included in the evaluation.

7. Conclusion

In conclusion, VR can offer an exciting and engaging ways to learn and teach. But there are still challenges that need to be overcome to make it feasible for schools. We wish to offer some recommendations on how to create an engaging VR learning experience, from which both students and teachers will benefit. Our core findings were that an immersive experience motivates students and encourages them to learn more. Moreover, time and costs are still crucial factors whether the VR laboratory will be included into school routines or not. Additionally, VR offers many possibilities and improvements for the knowledge gaining process, such as enabling the performance of dangerous experiments in a safe way or visualizing invisible processes.

Although participants rated the interaction with the laboratory as good, there is still potential for improvement, especially in the design and usability. Both pupils and student teachers requested a more accessible method for easier classroom integration. Future research should consider the potential of mobile VR devices such as smartphones or standalone devices and explore how we can integrate them into classrooms. For future work, it would be useful to port the current framework to mobile devices and explore how we can integrate these in classrooms. An important step would be to discuss the current results with experienced teachers and to obtain their feedback on the identified challenges and recommendations. Future studies could then investigate the different use cases to develop a pedagogical model for schools. In future the effect of not wearing earphones and of receiving tips, breaks and immersion could be investigated, or also if it could be a useful setup to have one student performing an experiment while one or more peers give hints on how to complete the tasks involved.

Acknowledgement

At this point, we would like to thank all participants of this study and all those people who are and were involved in the development process. This work was partly funded by the Ministry of Education, Science and Research of Austria.

References

Abdelaziz, M. A., Alaa El Din, M., & Senousy, M. B. (2014). Challenges and issues in building virtual reality-based e-learning system. *International Journal of e-Education, e-Business, e-Management and e-Learning, 4*(4), 320-328. doi: 10.7763/ijeeee.2014.V4.347

Abulrub, A. H. G., Attridge, A. N., & Williams, M. A. (2011). Virtual reality in engineering education: The future of creative learning. In *2011 IEEE global engineering education conference (EDUCON)* (pp. 751-757). doi:10.1109/EDUCON.2011.5773223

Asıksoy, G., & Islek, D. (2017). The Impact of the Virtual Laboratory on Students' Attitude in a General Physics Laboratory. *International Journal of Online Engineering, 13*(4), 20-28. doi: 10.3991/ijoe.v13i04.6811

Bogusevschi, D., Muntean, C., & Muntean, G. M. (2020). Teaching and learning physics using 3D virtual learning environment: A case study of combined virtual reality and virtual laboratory in secondary school. *Journal of Computers in Mathematics and Science Teaching, 39*(1), 5-18.

Brooke, J. (1996). SUS: A quick and dirty usability scale. In *Usability evaluation in industry* (pp. 189-194). London, UK: Taylor & Francis Ltd.

Callaghan, M. J., Eguíluz, A. G., McLaughlin, G., & McShane, N. (2015). Opportunities and challenges in virtual reality for remote and virtual laboratories. In *Proceedings of 2015 12th International Conference on Remote Engineering and Virtual Instrumentation* (pp. 235-237). doi:10.1109/REV.2015.7087298

Corter, J. E., Nickerson, J. V., Esche, S. K., Chassapis, C., Im, S., & Ma, J. (2007). Constructing reality: A Study of remote, hands-on, and simulated laboratories. *ACM Transactions on Computer-Human Interaction (TOCHI), 14*(2), 7-es. doi:10.1145/1275511.1275513

De Jong, T., & Van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of educational research, 68*(2), 179-201. doi: 10.3102/00346543068002179

De Jong, T., Linn, M., & Zacharia, Z. (2013). Physical and virtual laboratories in science and engineering education. *Science, American Association for the Advancement of Science*, *340(6130),* 305-308. doi: 10.1126/science.1230579

Dempsey, P. (2016). The teardown: HTC Vive VR headset. *Engineering & Technology, 11*(7-8), 80-81. doi: 10.1049/et.2016.0731

Dori, Y. J., Hult, E., Breslow, L., & Belcher, J. W. (2007). How much have they retained? Making unseen concepts seein in a freshman electromagnetism course at MIT. *Journal of Science Education and Technology, 16*(4), 299-323. doi: 10.1007/s10956-007-9051-9

Freedman, R. A. (1996). Challenges in teaching and learning introductory physics. In *from High-Temperature Superconductivity to Microminiature Refrigeration* (pp. 313-322). doi:10.1007/978-1-4613-0411-1_26

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. In *Proceedings of the National Academy of Sciences*, 111(23), 8410-8415. doi:10.1073/pnas.1319030111

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A Six-thousand-student survey of mechanics test data for introductory physics courses. *American journal of Physics, 66*(1), 64-74. doi:10.1119/1.18809

Ip, H. H., & Li, C. (2015). Virtual reality-based learning environments: Recent developments and ongoing challenges. In *International Conference on Hybrid Learning and Continuing Education* (pp. 3-14). doi:10.1007/978-3-319-20621-9_1

Kay, R., & Loverock, S. (2008). Assessing emotions related to learning new software: The computer emotion scale. *Computers in Human Behavior*, *24*(4), 1605-1623. doi:10.1016/j.chb.2007.06.002

Liou, W. K., & Chang, C. Y. (2018). Virtual reality classroom applied to science education. In *Proceedings of the 23rd International Scientific-Professional Conference on Information Technology (IT)* (pp. 1-4). doi:10.1109/SPIT.2018.8350861

Mayer, R. E. (2002). Cognitive theory and the design of multimedia instruction: An Example of the two-way street between cognition and instruction. *New directions for teaching and learning, 2002*(89), 55-71. doi:10.1002/tl.47

Moore, E. B., Chamberlain, J. M., Parson, R., & Perkins, K. K. (2014). PhET interactive simulations: Transformative tools for teaching chemistry. *Journal of Chemical Education, 91*(8), 1191-1197. doi:10.1021/ed4005084

Olson, S., & Riordan, D. G. (2012). Engage to Excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Report to the President. *Executive Office of the President.*

Pirker, J., Berger, S., Guetl, C., Belcher, J., & Bailey, P. H. (2012). Understanding physical concepts using an immersive virtual learning environment. In *Proceedings of the 2nd European Immersive Education Summit, Paris* (pp. 183-191).

Pirker, J., Gütl, C., Belcher, J. W., & Bailey, P. H. (2013). Design and evaluation of a learner-centric immersive virtual learning environment for physics education. In *International Conference on Human Factors in Computing and Informatics* (pp. 551 - 561). doi:10.1007/978-3-642-39062-3_34

Pirker, J., Lesjak, I., & Guetl, C. (2017a). Maroon VR: A Room-scale physics laboratory experience. In *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)* (pp. 482-484). doi:10.1109/ICALT.2017.92

Pirker, J., Holly, M. S., Hipp, P., König, C., Jeitler, D., & Gütl, C. (2017b). Improving physics education through different immersive and engaging laboratory setups. In *Interactive mobile communication, technologies and learning* (pp. 443-45). doi:10.1007/978-3-319-75175-7_44

Pirker, J., Lesjak, I., Parger, M., & Gütl, C. (2018). An educational physics laboratory in mobile versus room scale virtual reality-a comparative study. In *Online Engineering & Internet of Things* (pp. 1029-1043). doi:10.1007/978-3-319-64352-6_95

Pirker, J., Holly, M., Almer, H., Gütl, C., & Belcher, J. W. (2019a). Virtual reality STEM education from a teacher's perspective. In *Proceedings of 5th International Conference of the Immersive Learning Research Network.* doi:10.3217/978-3-85125-657-4-14

Pirker, J., Holly, M., Lesjak, I., Kopf, J., & Gütl, C. (2019b). MaroonVR—An Interactive and immersive virtual reality physics laboratory. In *Learning in a Digital World* (pp. 213-238). doi:10.1007/978-981-13-8265-9_11

Pirker, J., Holly, M., & Gütl, C. (2020). Room scale virtual reality physics education: Use cases for the classroom. In *Proceedings of the 6th International Conference of the Immersive Learning Research Network (iLRN)* (pp. 242-246). doi:10.23919/iLRN47897.2020.9155167

Reeve, J., Jang, H., Carrell, D., Jeon, S., & Barch, J. (2004). Enhancing students' engagement by increasing teachers' autonomy support. *Motivation and emotion, 28*(2), 147-169. doi:10.1023/B:MOEM.0000032312.95499.6f

Safikhani, S., Holly, M., & Pirker, J. (2020). Work-in-Progress—Conceptual Framework for User Interface in Virtual Reality. In *Proceedings of 6th International Conference of the Immersive Learning Research Network, iLRN 2020* (pp. 332-335). doi:10.23919/iLRN47897.2020.9155207

Sanders, M. (2008). Stem, stem education, stemmania. *The Technology Teacher* (pp. 20-26).

Slavova, Y., & Mu, M. (2018). A Comparative study of the learning outcomes and experience of VR in education. In *Proceedings of the Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 685-686). IEEE. doi:10.1109/VR.2018.8446486

Stark, L. W. (1995). How virtual reality works: Illusions of vision in "real" and virtual environments. In *Human Vision, Visual Processing, and Digital Display VI* (Vol. 2411, pp. 277-287). International Society for Optics and Photonics. doi:10.1117/12.207546

Velev, D., & Zlateva, P. (2017). Virtual reality challenges in education and training. *International Journal of Learning and Teaching, 3*(1), 33-37. doi:10.18178/ijlt.3.1.33-37

Wieman, C., & Perkins, K. (2005). Transforming physics education. *Physics Today, 58*(11), 36. doi: https://doi.org/10.1063/1.2155756

Zeidler, D. L. (2016). STEM education: A Deficit framework for the twenty first century? A Sociocultural socioscientific response. *Cultural Studies of Science Education, 11*(1), 11-26. doi:10.1007/s11422-014-9578-z

Lyons, L., & Mallavarapu, A. (2021). Collective Usability: Using Simulation Tools to Explore Embodied Design Challenges in Immersive, Shared Mixed-Reality Experiences. *Educational Technology & Society*, *24* (2), 120–135.

Collective Usability: Using Simulation Tools to Explore Embodied Design Challenges in Immersive, Shared Mixed-Reality Experiences

Leilah Lyons[1,2*] and Aditi Mallavarapu[2]
[1]New York Hall of Science, New York, USA // [2]University of Illinois at Chicago, Illinois, USA //
llyons@nysci.org // amalla5@uic.edu
[*]Corresponding author

ABSTRACT: In this paper we define the concept of collective usability, a complex systems perspective on usability that positions an entire group, not an individual, as the unit of analysis. Shared XR experiences have inherent temporal and spatial properties that produce emergent, collective impacts which can impede learners' engagement. Assembling large groups of users to test multiple design configurations is both logistically and financially impractical, however. We demonstrate the practical value of exploring the design space of an XR experience with a simple observation-informed Agent-Based Model. We used the model to explore how changes in the number of simultaneous users, and in the size, placement, and interaction duration of the proffered interactives, could affect collective access to a large-scale, mixed-reality, multi-user museum exhibit. (Collective access, an element of collective usability, is the degree to which users can gain access to each of the different interactives.) With this simple model, we explored (1) how the bottom-up propagation of individual-level design properties can affect collective outcomes, as when certain interactives' linger times cause a bottleneck, and (2) how the top-down propagation of collective design constraints can be used to guide individual-level design, as when we determined thresholds for the "stickiness" and "repeat allure" of an interactive to improve collective access. The final design of the exhibit implemented many of the design guidelines uncovered by the model. We argue that collective usability models could be useful for addressing a range of collective usability issues, beyond collective access, for temporally and spatially sensitive XR learning environments.

Keywords: Collective usability, Agent-based models, Usability methods, Informal learning, Shared XR

1. Introduction

Improved technologies are allowing designers to realize the dream of creating immersive, multi-user experiences akin to science fiction "holodecks." The embodied nature of XR experiences introduce new design challenges, as such systems are acutely sensitive to both temporal and spatial relationships between people and interactive objects or areas. We argue that in addition to traditional individual-level usability concerns, designers of multi-user XR experiences need to design for collective usability, a term we define as the degree to which a *group* of simultaneous users can make use of an interactive experience where the human-computer and human-human interactions combine to form a complex system. An individual user is the unit of analysis in traditional usability, whereas a group of users is the unit of analysis for collective usability. Where traditional usability gives primacy to the interaction between an individual human and a computer, collective usability embraces a complex systems perspective, wherein the individual interactions between humans and computers (and between humans and humans) combine to generate emergent effects that can only be measured at the group level. Attending to collective usability is important because designs with high individual usability (like a public transit ticket machine that ensures users pay the correct amount for their destination) may have inadvertently poor collective usability (e.g., a long line forms). Collective usability is especially critical for ensuring equitable learning environments, where all learners should get a chance to meaningfully engage with the proffered experiences.

The goal of this paper is to describe some of the challenges faced and the analysis techniques used when designing a large-scale Mixed Reality (MR) experience for a museum. For this study we videotaped 5 hours of more than 100 visitors' interactions with an early pilot exhibit, which consisted of six interactives that each supported a different type of full-body interaction (some gesture-based, some tangible-based). We used this data to empirically compute the statistical distributions for visitor engagement with these six interactives. We used these distributions to develop an Agent-Based Model (ABM) simulation to explore how changes in the number of simultaneous users and in the duration, size, and number of the proffered interactives could affect collective usability. We were specifically interested in a form of collective usability we dub collective access: the degree to which visitors could try each of the different interactives. We propose that collective usability models, of which our ABM is an example, could be used to represent a wide range of embodied usability challenges and thus become part of the toolkit of XR designers.

2. Background and prior work

2.1. Modeling users to support design processes

Using model-based simulations to support usability design and evaluation is not new — as early as the 1980s the GOMS model was used to engineer the usability of single-user interactives, by accounting for how human cognitive performance impacts the execution of software tasks (Card, Moran, & Newell, 1983). These "engineering models" should be "deliberately approximate … [including] just the level of detail necessary to do the design job … while keeping the modeling effort tractable" (John & Kieras, 1996, p. 4).

A number of researchers have explored how to model the behavior of users in collaborative scenarios to support the design process (Ivory & Hearst, 2001). For example, extending UML (a modeling language used by software developers) to describe group processes within design specifications (Garrido & Gea, 2001), or embedding the relational roles between group members into task models (Herrmann et al., 2004). The problem with many of these models is that, like GOMS, they assume a task model — i.e., that users will be engaged in specified tasks while working together. Unfortunately, there is often a gap between the intended tasks and how group interaction actually unfolds, suggesting models need to be based on how people actually act, not on idealized task execution sequences (Poltrock & Handel, 2010).

XR Learning Environments (LEs) tend to support open-ended use, so it is especially critical for collective usability models to be based on observations of how learners actually behave. We argue that for most XR LEs, minutely modeling learner cognition is not necessary for creating an engineering model. All that is needed is a minimal representation commonly seen behaviors, like the "foraging" behaviors of visitors in museums. It *is* important, however, to model elements of the shared domain context that can afford or constrain users' interactions with the system or their human companions (Wurdel, 2009). Because the interaction in XR LEs is stretched across a real or imagined physical setting, a modeling approach is needed that takes into account physical aspects like the spatial arrangement of interaction opportunities and other users.

2.2. Model-based methods in multi-user system design

Multi-user system researchers also use models to inform design, albeit in the form of highly formalized models used to validate competing system designs. They too have acknowledged the cyclical nature of the usability of multi-user systems, wherein "not only will the individual behavior of a user be affected by changes to the system through its collective use, but the system can also have an effect on the collective behavior of the users" (Massink et al., 2008). Researchers have predominantly used models to examine timing complications, making use of an array of formal models augmented with timing information and constraints — like finite automata, specialized algebras, and petri nets — to validate their designs (Bernardo & Corradini, 2004). Because most researchers were validating software like shared file management systems (e.g., Massink et al., 2008), the physical or spatial aspects of shared interaction were not modeled. One exception is a research group that represented spatial locations within a building to model a building's dynamic signage system (Harrison et al., 2008). They later used an ABM elaborated with force calculations to validate how the dynamic signage system performed against a static signage during evacuations (Langner & Kray, 2014).

Our work differs from this specific approach, and from the work on multi-user systems in general, in that we use an ABM as a *design tool* for collective usability. As a design tool, we employ ABM to explore the design space and to derive creative insights and guidelines to drive further design processes, not to formally validate designs.

2.3. Design challenges of full-body XR Learning Environments (LEs)

Museums and researchers exploring classrooms of the future are increasingly interested in the educational potential of sensing technologies that allow learners to use their bodies to interact with XR LEs. There are many rationales for using full-body movements to support learning activities. Full-body interactions can be designed to capitalize on *embodiment* by activating the proprioceptive aspects of cognition; they can take advantage of how cognition is *embedded* in the physical and social structures of the environment; or they can *extend* beyond individual organisms to support distributed cognition (Roberts & Lyons, 2017). One exhibit that relies on the *embodiment* aspect of whole-body interaction asks a visitor to control the path of a simulated meteor by running at different angles and speeds, using body sensations to build an understanding of force and motion (Lindgren, Tscholl, Wang, & Johnson, 2016). Many museums have developed camera-based XR interactives where visitors jump or dance, where usability designers need to design for how performances are *embedded* in the audience's spectator experience (Peltonen et al., 2008; Reeves, Benford, O'Malley, & Fraser, 2005). Sometimes *embedding* whole-body interactions within a spectator experience leads to learners coordinating and coaching one another, which *extends* the interaction into the realm of distributed cognition, as with an interactive slide (Malinverni, Ackermann, & Pares, 2016) or a classroom game about friction (Enyedy, Danish, & DeLiema, 2015).

The exhibit in this work employs full-body interactions to *extend* learner cognition to embrace shared management of a dynamic complex system. *Extended* full-body interaction design poses special usability challenges in the form of needing to manage potential interaction effects between different users. In prior work that *extended* cognition through whole body interactions, designers put the onus of managing interaction effects back on the learners — making learners responsible for monitoring one another and ensuring their actions don't negatively impact the collective experience. We argue that as XR experiences get more complex, designers need tools to examine the design challenges that arise from multiple users' actions, and to experiment with how different designs might impact those interaction effects.

2.4. Agent-based models

The review of prior work shows that there is a need for an "engineering model" (John & Kieras, 1996) to support design processes for XR LEs. Specifically, a modeling approach is needed that allows both spatial and temporal aspects of the design space to be represented. These requirements led us to use an ABM, defined as "(1) a computational method that enables a researcher to create, analyze, and (2) experiment with (3) models composed of (4) agents that interact within (5) the environment" (Gilbert, 2020). ABMs model a complex system by defining basic scenario properties (like the placement of food sources near an anthill) and the rule-based behaviors of independent agents (like the pheromone-governed movements of ants), to represent the complex behaviors that emerge (like coordinated food gathering). ABMs have been used to model a wide range of scientific and social collective activities, like traffic flows, military scenarios, movements of customers in stores, supply chain logistics, and administrative workflows (Abar et al., 2017). To the best of our knowledge, however, ABMs have never been promoted as a tool for usability design, apart from their use to design the behavior of artificially intelligent agents for ubiquitous computing (e.g., Mangina et al., 2010).

3. Design of pilot exhibit

The New York Hall of Science received a grant to develop a collaborative MR exhibit to teach visitors about sustainability and complex systems thinking in the context of a human-natural system. The exhibit needed to serve all ages, but the learning goals placed a special focus on middle school–age visitors. Complex systems thinking is challenging, and middle-school learners especially struggle with identifying the causal relationships between system components (Grotzer, 2012; Goldstone, 2006, Hmelo-Silver et al., 2007). One way to help learners develop a sense of a system's causal mechanisms is to have them take on first-person perspectives of different system agents (Jacobson & Wilensky, 2006). For this reason, it was decided to allow visitors to try out the roles of multiple agents within the system, allowing them to develop a firsthand appreciation for why, say, both farmers and city residents might draw down a water supply without realizing their actions could impact others. To encourage learners to try out multiple roles and thus broaden their understanding of the system, each role needed to be executed via an interactive that didn't require too much linger time. Moreover, because museums are free-choice learning environments, the operation of the interactive components could not be tightly coupled, i.e., *require* ongoing tight coordination between visitors (Salvador et al., 1996), although visitors could certainly choose to coordinate their actions. The design firm, DesignIO, created an initial functional prototype to explore some of these design issues.

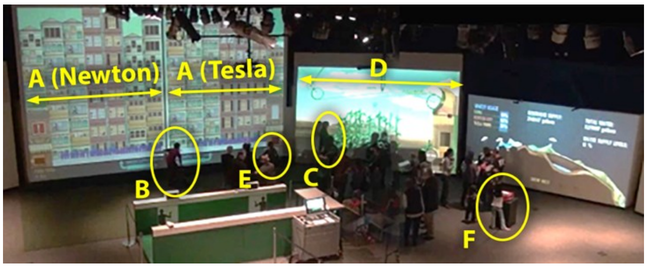


Figure 1. Pilot exhibit. Whole-body interactives are marked with letters and described in Table 1

The simulation consisted of two cities, Newton and Tesla (left screen in Figure 1), a farm where visitors could raise corn crops (middle screen in Figure 1) and a water-source selection screen (the screen on the far right in Figure 1). Water was shared among the farm and the two cities, meaning that if visitors didn't carefully manage urban growth there would be a water crisis, and subsequently a food crisis. The interactives available to individual users are illustrated in Figure 2 and described in Table 1.

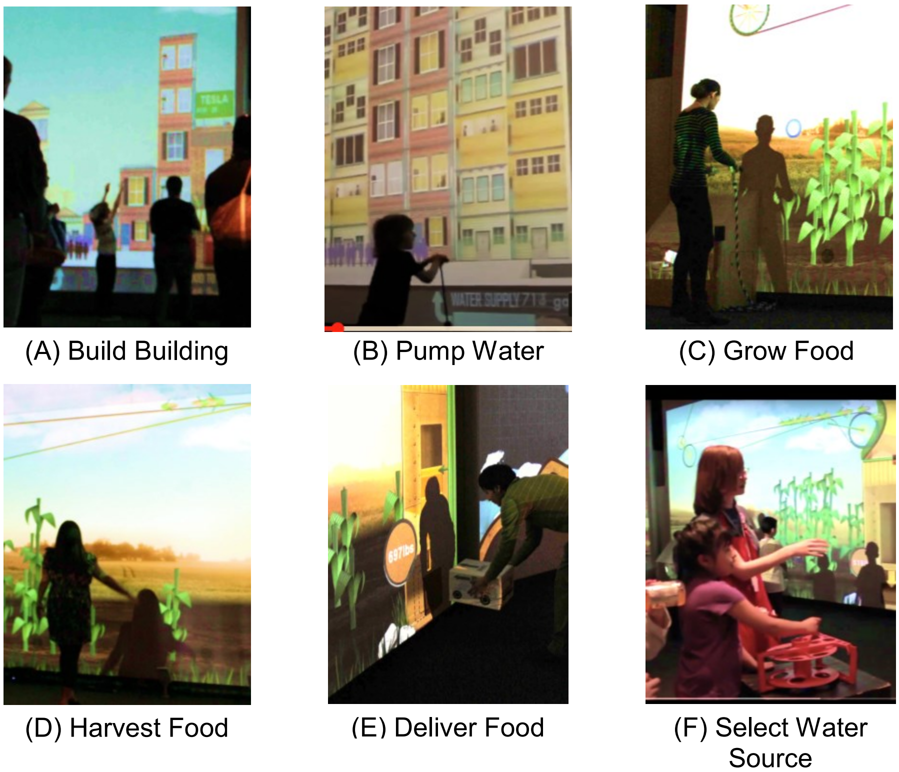


Figure 2. Snapshots of the six full-body interactives in the pilot exhibit

Table 1. Interactives available in the exhibit pilot

Action	Site of action	Max users	Interdependencies	Description of action enactment
(A) Build Building	Newton/ Tesla screen	7/7	Increases need for (E) Increases need for (B) Competes with other city	Visitor raises hands straight up above head to add another story to a building
(B) Pump Water	Pump near Newton/ Tesla screen	1	Affected by (F) Competes with (C)	Visitor operates a bicycle pump to pump water from the selected water source to both cities
(C) Grow Food	Hose in front of Farm screen	1	Affected by (F) Permits (D) Competes with (B)	Visitor uses a sprayer hose to irrigate farmlands (blue circle on screen indicates spray target), which sprouts corn
(D) Harvest Food	Farm screen	8	Requires (C) Permits (E)	Visitor uses a swiping arm motion to harvest mature corn, which gets deposited in silo
(E) Deliver Food	Between Farm and Newton/ Tesla screens	1	Requires (D) Choice in delivery city can induce competition	Visitor uses a cardboard "truck" box to pick up food from the Farm silo and deliver it to a depot in either Newton or Tesla
(F) Select Water Source	Rotary dial in front of Water Source screen	1	Permits (B) Permits (C) Is affected by (B) and (C)	Visitor rotates a dial, which sets the source of water used in the farm and the cities (e.g., aquifer, captured rainwater, river, etc.). When drained a new source must be chosen.

4. Methods

The purpose of this research was to explore how different design decisions might affect visitors' collective access. Collective access, the degree to which visitors could try each of the interactives, was important because allowing visitors to try on a breadth of different roles was theorized to support the exhibit's learning goal of encouraging systems thinking. To build such a model, we made the simplifying assumption to represent a visitor's engagement with just a single metric: linger time, based on distributions of how long real visitors would use a given interactive.

4.1. Participants

Participants were a mix of regular daily visitors to the museum who were told of the testing sessions at admission and/or saw signs outside the test gallery, and museum members who had received an email advertising the pilot testing sessions. The test sessions were held on several weekdays during spring break, so local families were the primary attendees. The apparent age ranges and cultural backgrounds of attendees were in keeping with typical visitorship for our institution, which tends to skew slightly young (average age of children being around 12) and which tends to be quite diverse. We used an implied consent procedure approved by our IRB, wherein a lollipop sign indicated that entering the test space constituted consent to being videotaped. Because collective usability takes the group as the unit of analysis, we did not collect information from individual visitors. We thus did not collect statistics on the number of visitors who used the exhibit, but there seemed to be about 20 visitors present at any given time, and both museum attendance data and the staff observations suggest that more than 100 individual visitors experienced the exhibit during our data collection.

4.2. Collecting visitor linger time statistics

The pilot exhibit was installed in a temporary gallery in the museum for several days, allowing us to videotape 4 hours and 58 minutes of visitors' interactions with the exhibit from a gallery on the rear wall (see Figure 1). Six research personnel coded the videos, noting when a visitor began using one of the interactives in Table 1, and when they stopped, capturing these start/stop events as timestamps. Using those timestamps, we then computed the average and standard deviation of visitor linger time (see Table 2). We did not compute inter-coder reliability, as the events were very obvious.

Table 2. Observations of visitors' interaction linger times

Interactive	Number of observations	Mean linger time, in seconds	Standard deviation
(F) Select water source	$n = 283$	$M = 7$	$SD = 9$
(A) Build building (Tesla)	$n = 103$	$M = 13$	$SD = 13$
(A) Build building (Newton)	$n = 83$	$M = 16$	$SD = 19$
(D) Harvest food	$n = 642$	$M = 20$	$SD = 28$
(C) Grow food	$n = 528$	$M = 21$	$SD = 33$
(B) Pump water	$n = 165$	$M = 42$	$SD = 53$
(E) Deliver food	$n = 132$	$M = 120$	$SD = 100$

These results show that the "stickiest" activity is food delivery (E) — visitors would run back and forth between the Farm silo and the Newton and Tesla food depot points for an average of 120 seconds before relinquishing the truck to another user. The least sticky activity is the water source selection wheel (F) — which makes sense, since the action is the closest one to being atomic (meaning that once the user chooses a water source, they are done, like clicking a button).

4.3. Simulating visitor use of exhibit interactives

The informal qualitative observations of the designers and researchers found that the pilot exhibit did not support collective access well, with numerous instances of people waiting around interactives to get a turn. This left a number of questions concerning how the exhibit could be modified to ensure that visitors could get a chance to try each of the interactives. We thus created an ABM to explore design variations virtually, without the cost and logistical overhead of revising and testing the prototype.

4.1.1. Building a simulation of the pilot exhibit in NetLogo

We selected NetLogo (Wilensky, 1999) to build our model. Originally developed for educational purposes, NetLogo has since been adopted by a number of professionals like urban planners for its convenience in quickly roughing out models of the intersection of people and places. We measured the screen sizes and angles and location of the tangible interactives to reproduce the pilot exhibit space (see Figure 3). For the tangible-based, single-user interactives (B, C, E, F), we set up rules to prevent other agents from using them if they were already in use. For the wall-based interactives (A, D), we used the maximum observed number of simultaneous visitors to determine the minimum wall space occupied per user, 1.875 feet, and used that to cap the number of simultaneous agents. Apart from (D), the interactives were designed to permit visitors to operate them regardless of their interdependencies, so we opted not to model the system's interdependencies. This simplification prevented us from going down the slippery slope towards fully modeling how visitors would reason about and respond to system interdependencies, which would be useful for modeling visitor learning, but not needed to inspect access issues. Such parsimony is recognized as good practice when constructing ABM (Gilbert, 2020).

4.1.2. Modeling agent behaviors

It is considered good practice to use the simplest possible representations of how agents behave to better explore emergent effects (Gilbert, 2020). We thus used Brownian motion as the simplest possible model of how visitors move within an exhibit space. With Brownian motion, agents are assigned a random trajectory they will follow until colliding with another agent or a physical object. Upon collision they get assigned a new random trajectory or (if the object is an available interactive) they will stop and interact. But as a check on the robustness of our results, we implemented two other plausible simplifications of how visitors could choose to move in the space: a Nearest model, where idle agents select as a target the nearest unoccupied interactive, and an Interest model, where idle agents select the unoccupied interactive that they have had the least experience with, with ties being broken by the nearest interactive. Regardless of the movement model, once an idle agent reaches an unoccupied interactive, they would occupy it for a duration of time drawn from the normal distribution for that interactive (gathered from observations of real users, see Table 2).

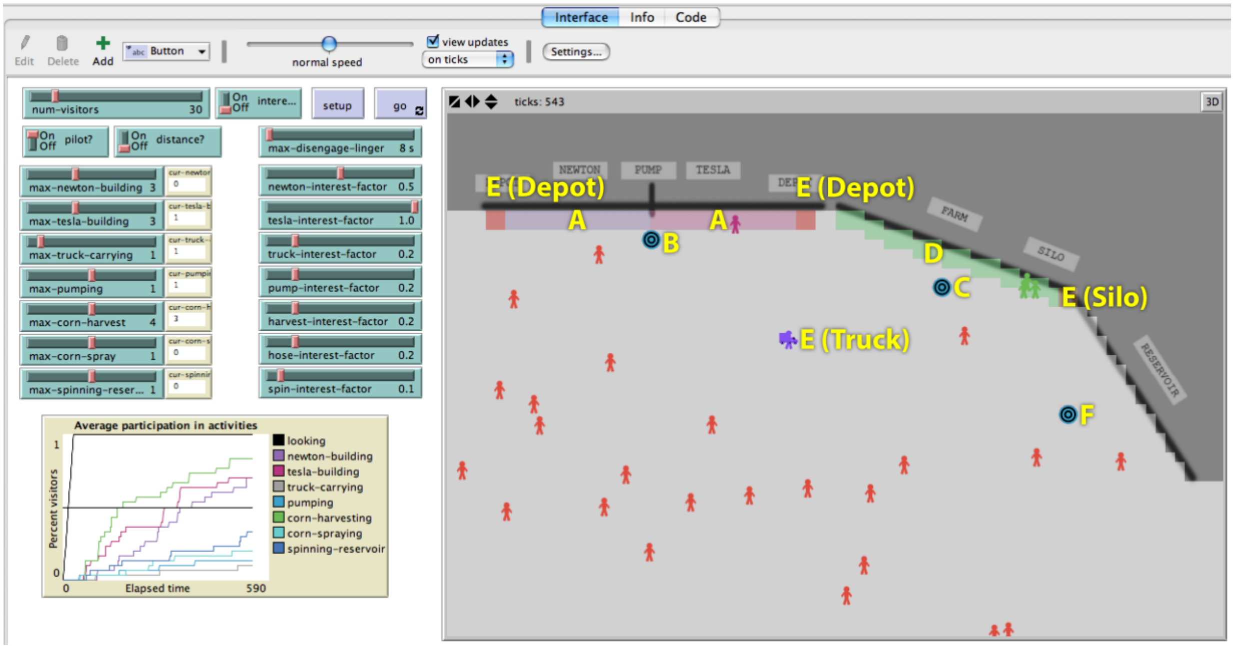


Figure 3. Agent-based simulation of visitors' participation in the Connected Worlds pilot. Colored squares mark where visitors can engage with an interactive

The Nearest movement model accounts for visitor "laziness" — the interaction space is quite large, and so visitors are likely to interact with a closer interactive if it is available. The Interest movement model represents the effect of novelty on visitor choices - all else being equal, visitors do prefer to try new or under-explored interactives. With the Interest model, for each agent, all interactives are initially assigned the same interest level of 1. The agent's interest level for each interactive is decreased by a fixed amount after using it. Visitors preferentially select available interactives with the highest remaining interest levels, with ties broken by selecting the closest interactive. This is a highly simplified analogue to more sophisticated value-cost models of visitor movements (Bitgood, 2006). We used the expert judgment of our research team, a mix of researchers and practitioners with extensive experience observing visitors, to select an interest decrement that produced plausible simulated visitor movements, eventually settling on a 10% decrease in interest. (Our choices in movement models, and implications for how much attention researchers should give to validating same, are tackled in the Conclusion).

5. Simulation results

The goal of this simulation-based exploration was to reduce the amount of time it would take for visitors to gain exposure to the majority of the exhibit interactives, i.e., to improve collective access. We thus used as our outcome the "half-used" metric: the time it would take at least 50% of the visitors to use each of the interactives. To generate half-used metrics we used a Monte Carlo approach. For each experimental configuration of the simulation model, we ran it 100 times, and averaged the results together. Each model would run until reaching the half-used state, or until 60 minutes of elapsed in-simulation time (60 minutes is well above the time visitors would be permitted to stay at the exhibit).

5.1. Baseline scenario under varied attendance

We first established a baseline simulation for 10, 20, 30, and 40 visitors to examine how the exhibit scaled with more users, and ran it under all three movement models to check for inconsistencies. Figure 4 shows that while the Nearest and Interest models do produce lower half-used times (which makes sense — there would be less aimless wandering), all three movement models produce linear regressions with similar slopes, suggesting that for each additional visitor, another 30-60 seconds of delay can be expected to be added to the half-used metric. This monotonic consistency is reassuring, suggesting that while in real life, visitors may exhibit a mix of movement characteristics (sometimes wandering, sometimes seizing a near interactive, sometimes seeking out a desired interactive), the collective access (as measured by the half-used metric) will fall within clear bounds.

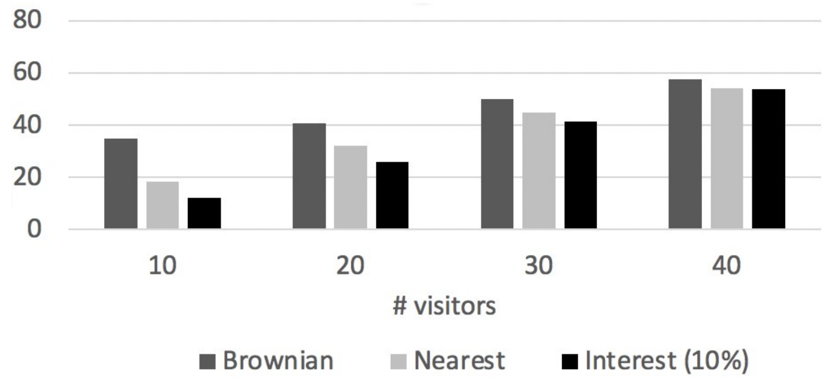


Figure 4. Time to reach the half-used state for all interactives under three different movement models, averaged across 100 Monte Carlo runs for each model and each number of visitors

5.2. Finding the source of half-used delays and extrapolating design variations to test

Our next step was to discover the major sources of delay — to determine this, we computed half-used metrics for each of the 7 possible interactives, under four different numbers of visitors (10, 20, 30, and 40) and the three different movement models (see Figure 5). The first thing to notice is that the overarching trends (in *magnitude* of half-used time, and in the *growth* in half-used times as the number of visitors increase) are consistent regardless of movement model, again suggesting that for our metric and problem space, the way we model visitor movement patterns is not a critical factor. The only major differences are that under the Interest model, the (F) Select Water Source time is reduced (because its low linger time allowed interested visitors easier access), and under the Nearest model, the (D) Harvest Food time is increased (because the farm is central to the exhibit, so it is often the nearest interactive and thus under more demand).

One strong pattern to notice is that the three screen-based activities that permit simultaneous users, (A) Newton Building, (A) Tesla Building, and (D) Harvest Food, do not show an increase in time to reach the half-used state, even as the number of visitors grows. At a collective level, it is this flat growth that we are seeking, as it shows that the interactives can flexibly scale to support larger groups. The screen sizes in our pilot were dictated by the gallery wall sizes, however, which were quite large — would we need to make the screens just as large in the final exhibit to preserve this scalability? This is a factor we wanted to explore further with our simulation.

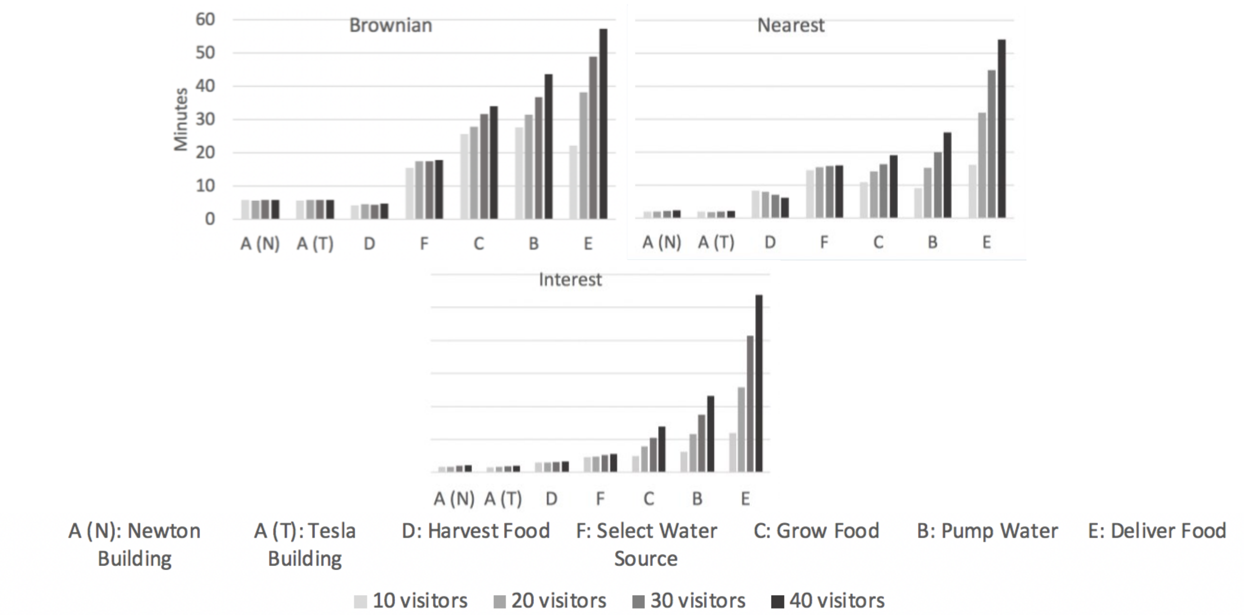


Figure 5. Time to reach the half-used state, in minutes, for each of the interactives as the number of visitors increases, under each of the movement models

The single-user tangible interactives, on the other hand, all show pronounced growth in half-used delay with increasing numbers of visitors, with one notable exception, the (F) Select Water Source activity. What differentiates the (F) from the other tangible interactives is its linger time (see Figure 6), which is closest to an atomic action (an instantaneous action-effect pairing, like a button click). This suggests that reducing the "stickiness" of individual tangible interactives (i.e., the amount of time a user will want to linger while operating it) is a design variation to explore for improving the collective user experience. However, it may be impossible to reduce the linger time of certain tangible interactives below a certain point — for example, carrying the food truck between the farm and the cities incurs a minimum linger time. Other alternatives to explore for improving collective access are replicating a tangible interactive with a long linger time (e.g., supplying multiple delivery truck interactives), or reducing how much a user would want to use an interactive again ("repeat allure").

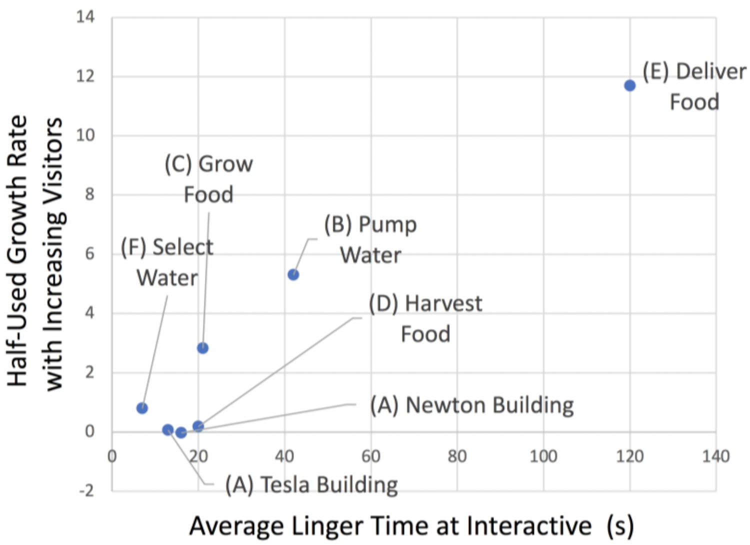


Figure 6. The relationship between the average linger time of each interactive, and the growth in half-used time as the number of visitors increases

5.3. Testing design variations with the ABM

The exploration of delay sources suggested that we should test four types of design variations for their impact on collective access: (1) changing the screen size for screen-based interactives, (2) decreasing the stickiness (i.e., linger time) of tangible interactives, (3) replicating tangible interactives, and (4) decreasing the "repeat allure" of the tangible interactives. We tested these parameters with all three movement models. All models produced similar results, so we present only the Nearest results for succinctness (the Nearest outcomes fell between the Brownian and Interest outcomes).

5.3.1. Manipulating screen width under varied attendance

The baseline results showed that the multi-user widescreen interactives were least impacted by growing numbers of visitors, so we wanted to see how important it was to use screens that were at least 13 feet wide. Would making the screens smaller hurt access, or would making them even larger benefit collective access? We wanted to hold as much else constant about our model as possible, including the relative positions of the interaction opportunities, so we opted only to manipulate the sizes of the Tesla Building and Newton Building screens (manipulating the farm screen width would have changed the entire layout). The Narrow Screen condition was half the size of the screen used in the baseline (6.5 feet), and the Wide Screen was 1.5 times the baseline (19.5 feet). We can see from Figure 7 that increasing the screen size does not appreciably impact the half-used time, while decreasing the screen size does remove its ability to scale with the number of visitors.

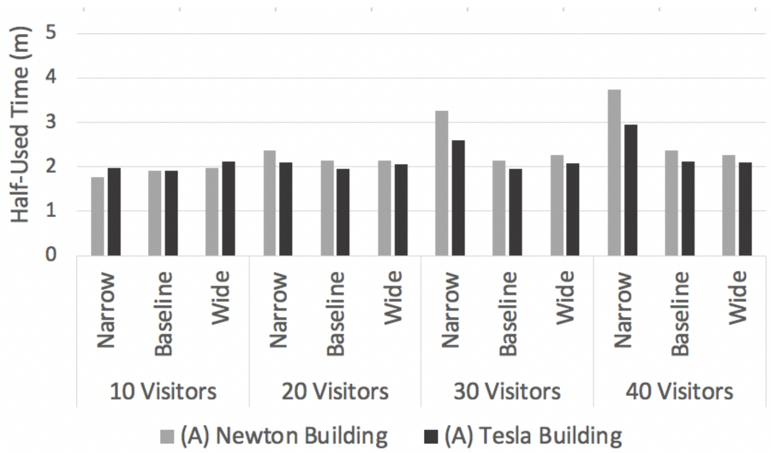


Figure 7. The impact of varying the screen size on the half-used time, in minutes, as the number of visitors increases

5.3.2. Reducing the "stickiness" of single-user tangible interactives

The baseline results showed that the single-user tangible interactives scaled least well when increasing the number of visitors, with the exception of the low-linger (F) Select Water Source wheel. Would decreasing the linger time of the other single-user interactives improve collective access? No designer wants to intentionally decrease the pleasurability of an interaction design, but it may be the case that the collective good demands it. There is research evidence that when visitors get highly engaged in single-user interactions within a shared exhibit, it negatively impacts the group experience (Lyons, 2009).

We reduced the linger time to a fixed 5 seconds (approximately the time needed to approach and execute an atomic action like a button click) to see how low stickiness would impact the half-used time. We implemented atomic versions of the (C) hose and the (B) pump (we did not test an atomic version of the (E) truck because its linger time depends on the distance between the farm's silo and the food delivery depots at Newton and Tesla). Figure 8 shows that if both interactives could be made less sticky, we could attain the flat growth with increased visitors as was seen for the (F) wheel. The (B) Pump Water interactive benefitted the most from a conversion to an atomic action — likely because its less-central location made it less well-trafficked, and thus more available for visitors.

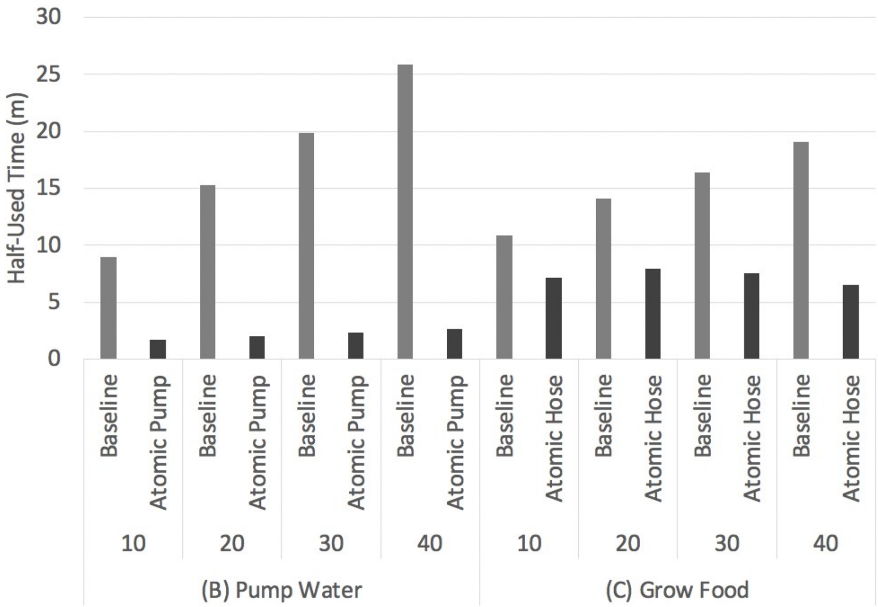


Figure 8. The impact on the half-used time, in minutes, of assigning a tangible interactive to have a fixed, atomic amount of linger time, as the number of visitors increases

5.3.3. The effect of replicating single-user tangible interactives

We replicated the three interactives with the largest growths in half-used time (B, C, and E) and present the results in Figure 9. An extra hose reduces the growth in (C) half-used time to be nearly flat, although the magnitude of reduction was less than what was seen when the hose was converted to an atomic action (Figure 8). While the rate of growth in half-used time remains high for the replicated (B) water pump and (E) truck, the *magnitude* of the half-used times decreased by around 50% as compared to baseline, regardless of visitor count. For tangible interactives with long linger times, it may be preferable to duplicate them than to remove stickiness, as removing stickiness may well involve removing the pleasurable aspects of interaction. However, a single replication may not be enough to meaningfully impact collective access — for example, the half-used time triples from a 10-visitor replicated truck scenario to a 40-visitor replicated truck scenario. This raises the question: could more replicated trucks attain a half-used measure that doesn't vary with increased numbers of visitors?

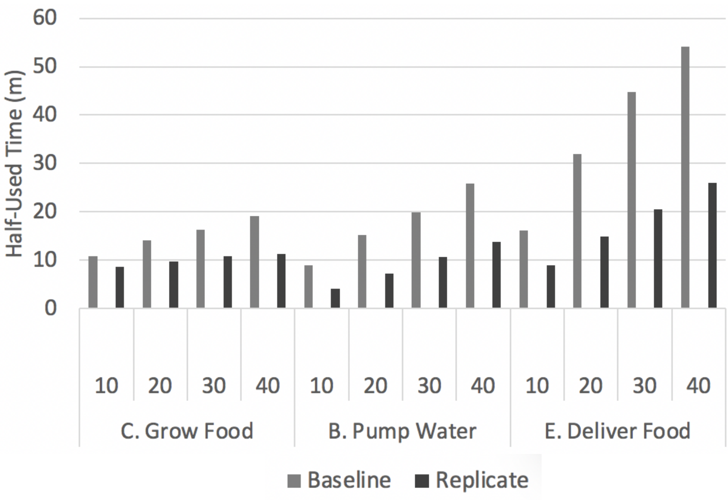


Figure 9. The impact on the half-used time, in minutes, of replicating the tangible interactives as the number of visitors increases

The results show that up to 8 trucks need to be present before the (E) ceases to account for more than 50% of the half-used time for the exhibit as a whole (see Figure 10). The half-used growth rate still doesn't flatten, however. Could it be that because the trucks are located in a highly-trafficked area, the same visitors end up picking up the trucks again and again, making it harder for other visitors to use them? The next section explores making that kind of repeat-use less appealing.

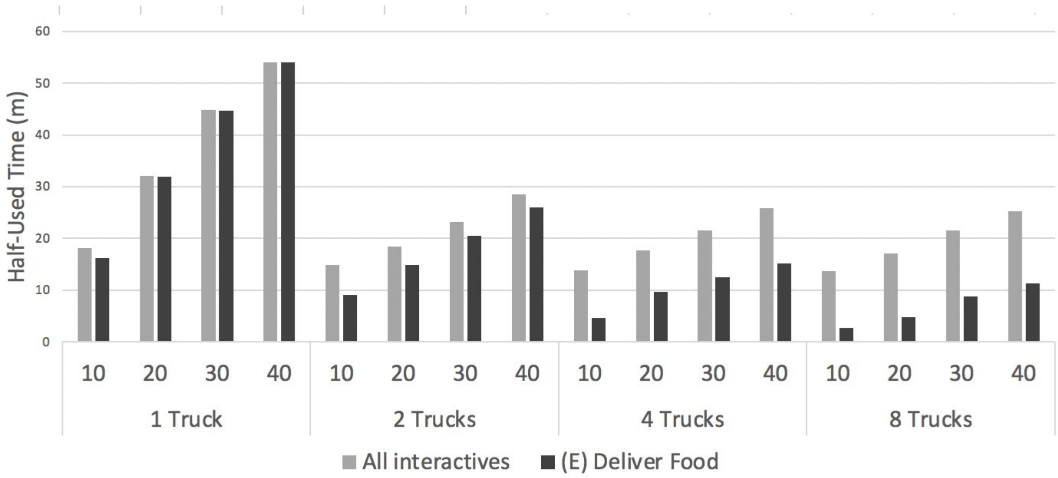


Figure 10. Time to reach the half-used state for all interactives and for the (E) Deliver Food interactive, in minutes, as the number of trucks and the number of visitors increases

5.3.4. Reducing the "repeat allure" of food delivery

Much like reducing "stickiness" by making interactives less engaging, designers can also make it less likely for users to opt to return to an interactive. Designers do not need to make the experience overtly unpleasant, but can instead think about ways to build in gentle impediments to make users less likely to immediately re-engage. With tangible user interfaces there are an array of ways of negatively manipulating their affordances, from literally adding friction (like making a wheel hard to turn) to adding weight (visitors might avoid picking up a heavy object again unless necessary).

We simulated a "repeat allure" decrease by manipulating the Interest-based movement model. In addition to the 10% drop used as the default for the Interest model, we experimented with three different degrees of decreasing the repeat allure for the Truck — a 25% drop, a 50% drop, and an extreme 100% drop in interest, compared to the Nearest baseline (see Figure 11). Decreasing the repeat allure does decrease the half-used magnitude, but still does not flatten the growth in half-used time. This suggests that while designers can incorporate negative affordances into the design of interaction opportunities that have unavoidably long linger times (e.g., tangible interactives that need to be carried from place to place), this alone may not be enough to scale with increased users.

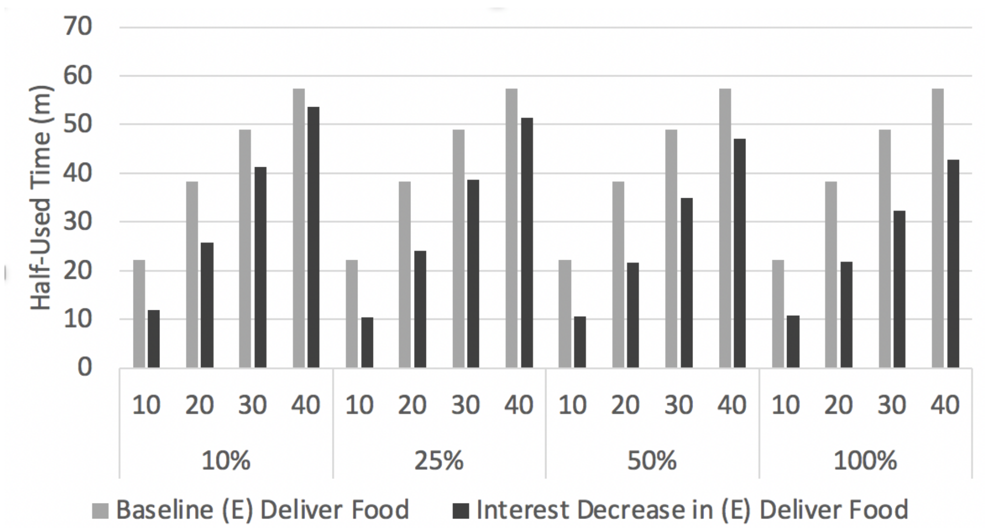


Figure 11. Time, in minutes, to reach the half-used state for the Food Delivery interaction type when visitor interest in using the Truck drops by growing percentages, as the number of visitors increases

5.4. Putting it all together — the kitchen sink simulation

One value of using ABMs for exploring design spaces is that they can reveal nonlinearities when multiple design variations are combined. In other words, sometimes combining two seemingly desirable changes can result in an unpredictably bad outcome. The final step in our analysis was thus to combine the best-result recommendations from each of the prior design variations into a single model and test them together (a "kitchen sink" simulation). The kitchen sink simulation combined the baseline screen width for (A) (we saw issues with decreasing the width, but no improvements with increasing width); the atomic (5 seconds linger time) versions of the (B) pump and (C) hose; and eight replicated (E) trucks with a 25% "repeat allure" decrease (larger repeat allure decreases did not dramatically reduce the half-used time, so we wanted to be conservative). The results, shown in Figure 12, reveal no negative interaction effects. To the contrary, combining these design recommendations finally flattened the growth in half-used rate. The magnitude of the decrease in exhibit-wide half-used time is also notable — a tenfold decrease from baseline when 40 visitors are present.

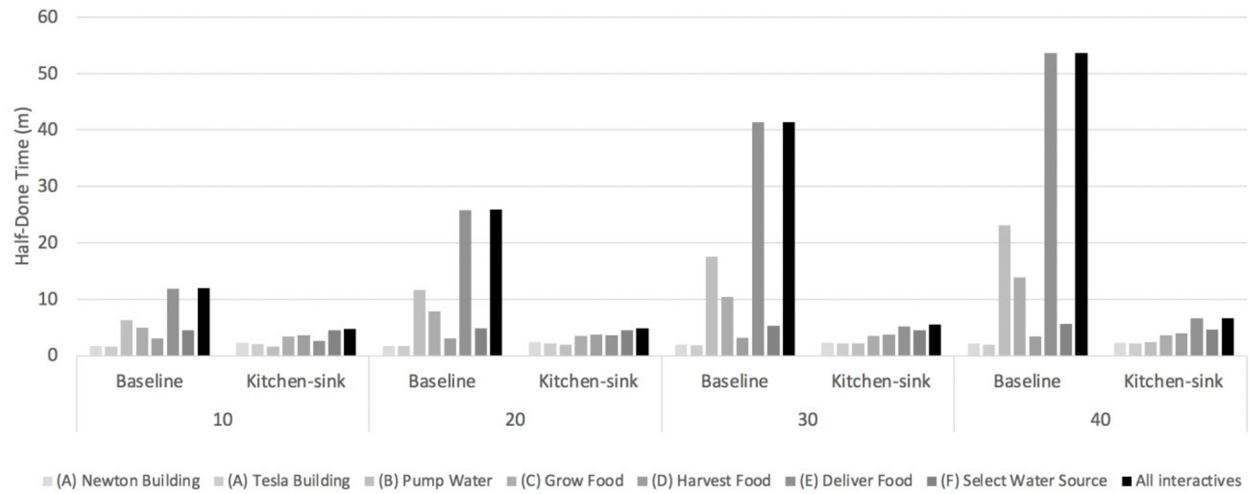


Figure 12. Time to reach the half-used state, in minutes, as the number of visitors increases, when all design recommendations are used (the "kitchen sink") compared against the baseline, using the Interest movement model

6. Discussion

The results show that ABM-based collective usability models can be used to profitably explore the impact of individual design changes on collective usability in immersive, shared XR LEs. It is worth emphasizing that while most laypeople assume all simulations are used to make precise predictions for specific outcomes, ABMs are uniquely useful for exploring the space of possible outcomes, and relative trends in those outcomes (Gilbert, 2020). Collective usability models are *thinking tools* for exploring a design space more effectively and efficiently. We demonstrated above how a collective usability model can be used to generate ideas for further model-based investigation. Below we describe the guidelines our analysis generated, and to what extent these guidelines were represented in the final exhibit design.

6.1. Simulation-derived collective design guidelines for MR exhibit

ABMs capture properties of complex systems like emergent effects, and thus can be used to generate design guidelines in two directions: bottom-up propagation of effects, and top-down propagation of design constraints.

6.1.1. Guideline derived from bottom-up effect propagation

With our model, we studied the bottom-up propagation of the linger times of individual interactives to measure the impact on collective access in terms of the half-used time. One might expect the interactives with longer linger times to impede collective access more, but we only found this to be true for the single-user tangible interactives. The multi-user screen-based interactives showed no problems scaling to support more users (see Figure 6). This suggests the design guideline:

DG1. When possible, situate interactive experiences on large, shared screens.

6.1.2. Guidelines derived from top-down propagation of design constraints

What we mean by “top-down propagation” is that we use the emergent outcome (the half-used time) to help set bounds on the design properties for individual interactives. Our simulation trials revealed several design guidelines applicable to our learning environment:

DG2. Keep the width of shared screen-based interactives to be at least 13 feet.
DG3. Reduce the “stickiness” of single-user interactives to reduce the linger time to be closer to an atomic button click in duration.
DG4. Replicate single-user tangibles that have irreducible linger times to avoid access issues. More copies than one might think (>=8) may be needed to prevent the interactives from being the limiting factor on collective access.
DG5. Mild friction may be introduced to the operation of single-user tangibles with irreducible linger times to reduce repeat allure.

6.2. Implementing the collective design guidelines in practice

Moving from prototype to final implementation can be a messy, nonlinear, and creative process. Collective usability design guidelines cannot and should not be the final word on how to design an immersive, shared XR experience, as collective usability is just one aspect of making such experiences successful. Designers need to balance collective usability against more intangible factors like the charm and pleasure that an individual interactive design can bring. While the final implementation of the exhibit did not follow all guidelines to the letter, they are present in various ways.

Figure 13. Photograph of the final exhibit. A six-story simulated waterfall (center), a river valley (middle left) and a reservoir (middle right) deliver water to the exhibit floor, which visitors can divert to four different simulated biomes projected on the exhibit walls

The final exhibit’s learning goals are the same, but rather than maintaining two cities, visitors instead maintain the diversity of four different biomes: the Desert, the Grasslands, the Jungle, and the Wetlands (see Figure 13). These four biomes are each visualized on 22-feet wide projection screens (which far exceeds DG2’s recommendation, but which adds a “wow” factor to the experience). Rather than positioning single-user tangible interactives in front of screens to support interactions, the design situates the interactions within the screens themselves as much as possible via gestures, per DG1. Visitors can plant seeds in these four biomes by holding their hands up in front of the screens (see Figure 14) and letting their hands quickly drop, and can prune dead plants by swiping their arms (the same action used for (D) in the pilot). These interactions are very short in duration, in keeping with DG3. Visitors supply water to the biomes by dragging large (5-feet-long) stuffed “logs” covered in an IR-reflective fabric around the floor of the exhibit (see Figure 14), diverting the flow of water from the sources. The logs are the only single-user tangible interactives in the exhibit, but manipulating the positions of these logs can take time, as they need to be moved to different spatial locations. There are 10 logs in the exhibit, per DG4, and the logs are a bit heavy, which encourages visitors to move them only when needed, per DG5. We should note that this approach did not decrease visitors’ apparent enjoyment of using the tangibles — in fact, they seemed to find the difficulty part of the fun, even as it served its purpose of reducing the “repeat allure.” We propose that after decades of trying to make user interfaces more seamless and more efficient, interface designers may have something of a blind spot for how strategic friction can be charming, and should embrace creative approaches for reducing repeat allure when it is demanded by collective usability.

Figure 14. A visitor selecting which available seed to plant in the Grasslands biome

We did not formally assess the collective access of the final exhibit because the earlier problems with access were completely resolved — even when used in 15 minute sessions with school groups, no access problems were observed. Visitors could and did have the opportunity to try all of the different interactives within minutes. As a whole, the exhibit has been a great success, winning multiple national awards like the 2016 Jackson Hole Science Media Award and the 2017 American Alliance of Museums MUSE Gold Award.

7. Conclusions and future work

This paper makes the case that shared XR LE design requires attention to the "holistic" configuration of people and objects in time and space (Wurdel, 2009). It defines the concept of collective usability, which positions a group of users, rather than an individual user, as the unit of analysis as they make use of an interactive experience where the human-computer and human-human interactions combine to form a complex system. It further provides an initial proof-of-concept that ABMs, which can capture the temporal and spatial relationships in collective use scenarios, can be used as collective usability models to explore the design space for shared XR experiences. Using ABMs to explore the usability design space is another novel contribution of this work. Unlike the abstruse formal models used for multi-user system validation, ABMs are easily constructed to be "deliberately approximate" (John & Kieras, 1996), extending the reach of this approach beyond experts.

This is just a proof-of-concept, so many directions for future work remain. Collective access is just one of many collective usability metrics that designers may wish to explore. For example, collective awareness, a measure of the extent to which users have detected a stimuli, may be highly relevant to collaborative XR LEs because shared visual grounding is known to be critical to supporting collaboration (Gergle, Rosé, & Kraut, 2007). Architecture researchers have long known how sight lines can impact how people behave in shared spaces, and have developed ways to model and simulate sight lines that could be borrowed from to construct models of collective awareness (Kaynar, 2005).

The highly simplified visitor movements in our collective usability model were adequate for our investigation, but other problem domains (e.g., studying if an exhibition layout poses challenges for maintaining a 6-foot social distance) may need more sophisticated models. Fortunately, there is ample literature on how visitors move in museums, from how different social groups distance themselves (vom Lehn, Heath, & Hindmarsh, 2001) to Bitgood's value model of visitor circulation (Bitgood, 2006), to documentation of how different kinds of visitors move differently — children, especially, are more prone to prioritizing exploration over exploitation (Carlisle, 1985). Creating and validating a movement model that can reproduce the spatial paths of learners may be time intensive, but once created, it can be reused and shared.

We argue that collective usability models should be thought of as another design tool, like design games (Brandt & Messeter, 2004), or design cards (Hornecker, 2010), that help designers attend to an aspect of the design space that might be otherwise overlooked. Our proof-of-concept work demonstrates how these collective usability design tools can be used. An initial collective usability model is created, either using empirical data from a prototype tested with representative users, or using theory-grounded models of user behavior. Designers use the model to generate unanticipated questions about the design space, questions that can be systematically explored by altering the design parameters embedded in the model. As the trends in outcomes become clear, design guidelines that speak to the collective usability experience can be generated using bottom-up propagation of effects and top-down propagation of design constraints. Collective usability modeling is much more logistically and financially feasible for exploring the design space for XR experiences than repeated prototyping, and can help generate design questions (like reducing interactive stickiness and repeat allure) that designers (who are not accustomed to "detuning" interactives) may not think to explore otherwise.

Acknowledgements

This material is based upon work supported by the National Science Foundation under grant nos. 1123832, 1623094, and 1822864.

References

Abar, S., Theodoropoulos, G. K., Lemarinier, P., & O'Hare, G. M. P. (2017). Agent based modelling and simulation tools: A Review of the state-of-art software. *Computer Science Review, 24*, 13-33. doi:10.1016/j.cosrev.2017.03.001

Bitgood, S. (2006). An Analysis of visitor circulation: Movement patterns and the general value principle. *Curator: The Museum Journal, 49(*4), 463–475. doi:10.1111/j.2151-6952.2006.tb00237.x

Bernardo, M., & Corradini, F. (Eds.). (2004). *Lecture notes in computer science: Vol. 3185. Formal methods for the design of real-time systems*. Springer. doi:10.1007/b110123

Brandt, E., & Messeter, J. (2004). Facilitating collaboration through design games. In *Proceedings of the Eighth Conference on Participatory Design: Artful Integration: Interweaving Media, Materials and Practices* (pp. 121–131). doi:10.1145/1011870.1011885

Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum.

Carlisle, R. W. (1985). What do school children do at a science center? *Curator, 28*(1). doi:10.1111/j.2151-6952.1985.tb01687.x

Enyedy, N., Danish, J. A., & DeLiema, D. (2015). Constructing liminal blends in a collaborative augmented-reality learning environment. *International Journal of Computer-Supported Collaborative Learning, 10*(1), 7–34. doi:10.1007/s11412-015-9207-1

Garrido, J. L., & Gea, M. (2001). Modelling dynamic group behaviours. In Johnson, C. (Ed.), *Lecture Notes in Computer Science: Vol. 2220. Interactive Systems: Design, Specification, and Verification* (pp.128–143). doi:10.1007/3-540-45522-1_8

Gergle, D., Rosé, C. P., & Kraut, R. E. (2007). Modeling the impact of shared visual information on collaborative reference. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1543–1552). doi:10.1145/1240624.1240858

Gilbert, N. (2020). *Agent-based models* (2nd ed.). Thousand Oaks, CA: Sage.

Goldstone, R. L. (2006). The complex systems see-change in education. *Journal of the Learning Sciences*, *15*(1), 35–43. doi:10.1207/s15327809jls1501_5

Grotzer, T. (2012). *Learning causality in a complex world: Understanding of consequence*. New York, NY: Rowman & Littlefield Publishers, Inc.

Harrison, M. D., Kray, C., & Campos, J. C. (2008). Exploring an option space to engineer a ubiquitous computing system. *Electronic Notes in Theoretical Computer Science, 208*, 41–55. doi:10.1016/j.entcs.2008.03.106

Herrmann, T., Hoffmann, M., Kunau, G., & Loser, K. U. (2004). A Modelling method for the development of groupware applications as socio-technical systems. *Behaviour and Information Technology*, *23*(2), 119–135. doi:10.1080/01449290310001644840

Hmelo-Silver, C. E., Marathe, S., & Liu, L. (2007). Fish swim, rocks sit, and lungs breathe: Expert-novice understanding of complex systems. *Journal of the Learning Sciences*, *16*(3), 307–331. doi:10.1080/10508400701413401

Hornecker, E. (2010). Creative idea exploration within the structure of a guiding framework: The Card brainstorming game. In *Proceedings of the 4th International Conference on Tangible, Embedded, and Embodied Interaction* (pp. 101–108). doi:10.1145/1709886.1709905

Ivory, M. Y., & Hearst, M. A. (2001). The State of the art in automating usability evaluation of user interfaces. *ACM Computing Surveys, 33*(4), 470–516. doi:10.1145/503112.503114

Jacobson, M. J., & Wilensky, U. (2006). Complex systems in education: Scientific and educational importance and implications for the learning sciences. *Journal of the Learning Sciences*, *15*(1), 11–34. doi:10.1207/s15327809jls1501_4

John, B. E., & Kieras, D. E. (1996). The GOMS family of user interface analysis techniques. *ACM Transactions on Computer-Human Interaction (TOCHI), 3*(4), 320–351. doi:10.1145/235833.236054

Kaynar, I. (2005). Visibility, movement paths and preferences in open plan museums: An Observational and descriptive study of the Ann Arbor Hands-on Museum. In *Proceedings of the 5th International Space Syntax Symposium*, (pp. 1–15). Retrieved from http://kaynar-rohloff.com/papers/kaynar_SpaceSyntax05.pdf

Langner, N., & Kray, C. (2014). Assessing the impact of dynamic public signage on mass evacuation. In *Proceedings of the 3rd International Symposium on Pervasive Displays* (pp. 136–141). doi:10.1145/2611009.2611033

Lindgren, R., Tscholl, M., Wang, S., & Johnson, E. (2016). Enhancing learning and engagement through embodied interaction within a mixed reality simulation. *Computers and Education, 95*, 174–187. doi:10.1016/j.compedu.2016.01.001

Lyons, L. (2009). Designing opportunistic user interfaces to support a collaborative museum exhibit. In *Proceedings of the 9th International Conference on Computer Supported Collaborative Learning* (pp. 375–384). doi:10.3115/1600053.1600110

Malinverni, L., Ackermann, E., & Pares, N. (2016). Experience as an object to think with. In *Proceedings of the 10th International Conference on Tangible, Embedded, and Embodied Interaction* (pp. 332–339). doi:10.1145/2839462.2839477

Mangina, E., Carbo, J., & Molina, J. (Eds.). (2010). *Agent-based ubiquitous computing*. Amsterdam, The Netherlands: Atlantis Press.

Massink, M., Latella, D., Beek, M. H., Harrison, M. D., & Loreti, M. (2008). A Fluid flow approach to usability analysis of multi-user systems. In P. Forbrig & F. Paternò (Eds.), *Lecture Notes in Computer Science: Vol. 5247. Engineering Interactive System* (pp. 166–180). doi:10.1007/978-3-540-85992-5_15

Peltonen, P., Kurvinen, E., Salovaara, A., Jacucci, G., Ilmonen, T., Evans, J., Oulasvirta, A., & Saarikko, P. (2008). It's mine, don't touch!: Interactions at a large multi-touch display in a city centre. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1285–1294). doi:10.1145/1357054.1357255

Poltrock, S., & Handel, M. (2010). Models of collaboration as the foundation for collaboration technologies. *Journal of Management Information Systems, 27*(1), 97–122. doi:10.2753/MIS0742-1222270105

Reeves, S., Benford, S., O'Malley, C., & Fraser, M. (2005). Designing the spectator experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 741–750). doi:10.1145/1054972.1055074

Roberts, J., & Lyons, L. (2017). The Value of learning talk: Applying a novel dialogue scoring method to inform interaction design in an open-ended, embodied museum exhibit. *International Journal of Computer-Supported Collaborative Learning, 12*(4), 343–376. doi:10.1007/s11412-017-9262-x

Salvador, T., Scholtz J., & Larson J. (1996). The Denver model for groupware design. *SIGCHI Bulletin, 28*(1), 52–58.

vom Lehn, D., Heath, C., & Hindmarsh, J. (2001). Conduct and collaboration in museums and galleries. *Symbolic Interaction*, *24*(2), 189–216.

Wilensky, U. (1999). *NetLogo*. [Computer software]. Retrieved from http://ccl.northwestern.edu/netlogo/

Wurdel, M. (2009). Towards an holistic understanding of tasks, objects and location in collaborative environments. In M. Kurosu (Ed.), *Lecture Notes in Computer Science*: Vol. *5619. Human Centered Design* (pp. 357–366). doi:10.1007/978-3-642-02806-9_41

Chaker, R. Gallot, M., Binay, M., & Hoyek, N. (2021). User Experience of a 3D Interactive Human Anatomy Learning Tool. *Educational Technology & Society*, *24* (2), 136–150.

User Experience of a 3D Interactive Human Anatomy Learning Tool

Rawad Chaker[1], Mélanie Gallot[2], Marion Binay[2] and Nady Hoyek[2*]
[1]Université Lumière Lyon 2, ISPEF, ECP, Lyon Cedex 07, France // [2]Université de Lyon, Université Claude Bernard, Lyon 1. Inter-university Laboratory of Human Motor Performance (LIBM - EA 7424), Villeurbanne Cedex, France // rawad.chaker@univ-lyon2.fr // melanie.gallot@univ-lyon1.fr // marion.binay@univ-lyon1.fr // nady.hoyek@univ-lyon1.fr
[*]Corresponding author

ABSTRACT: Embodiment is particularly relevant for learning anatomy as the knowledge to be acquired is related to the body itself. Several tools using three-dimensional (3D) anatomical structures and avatars (e.g. augmented reality; virtual reality; immersive anatomy; 3D animations) were developed to enrich students' experience by including gestures and body movements into learning anatomy. We developed a new interactive 3D tool that allows personal body experience and enhances spatial representation of musculoskeletal functional anatomy. Students can analyze and recreate a series of movements in real-time 3D interactive settings. This paper shows our research and development approach. Following the development of our anatomy tool, we conducted a pilot and one experiment. The pilot study aimed at evaluating users' experience (UX) of our first prototype. Experiment I aimed at evaluating the UX of the second version of the tool two times in a pretest-training-posttest design. Students' spatial and motor imagery abilities as well as anatomy examination results were also collected. Our results provided evidence of UX enhancement. Accordingly students appreciated mainly the tool's hedonic (enjoyment) qualities. Overall, significant interactions were observed between students' UX, anatomy scores and motor imagery abilities. Finally, students' mental rotation ability predicted the increase of anatomy score. Cognitive sub-processes underlying functional human anatomy learning as well as students' identification through the avatar are discussed.

Keywords: 3D tool, Anatomy, User experience, UX, Spatial ability, Motor imagery

1. Introduction

Computing technologies have transformed anatomical sciences education during the last decade (Trelease, 2016). Numerous programs worldwide have integrated online digital learning tools as supplementary resources: e.g. eBooks (Pickering, 2015), social media (Pickering and Bickerdike 2016), massive open online courses (Swinnerton et al., 2017), 3D animations (Hoyek et al. 2014), smartphone and tablet applications (Lewis et al., 2014), 3D-printed specimens (McMenamin et al., 2014; Lim et al., 2016), augmented, mixed and virtual reality (Küçük et al., 2016; Moro et al., 2017).

The results of the studies that have investigated the impact of such digital resources on learning outcomes have proven to be variable. For instance, Khalil et al. (2005) did not report significant differences in learning outcomes when comparing computer-based interactive and paper-based static instructional materials. Several other studies did not find any beneficial pedagogical effects of 3D stereoscopic models or videos on anatomy learning outcomes (Saxena et al., 2008; Hopkins et al., 2011; Tan et al., 2012). Conversely, Nicholson et al. (2006) as well as Abid et al. (2007) reported that 3D-computer-based anatomy graphics enhanced medical students' learning outcomes. More recently, Hoyek et al. (2014) demonstrated the effectiveness of 3D digital animation compared to 2D drawings embedded into PowerPoint® slides in an authentic classroom context. Regarding X-reality or XR applications (augmented, virtual and mixed reality) there remains a paucity of robust empirical evidence to justify the efficiency of such resources on learning outcomes (Clunie et al., 2018). While Moro et al. (2017) found that learning outcomes remained unchanged and that students exhibited blurred vision headaches and dizziness, Küçük et al. (2016) found a positive effect of augmented reality on learning outcome as well as a reduction of cognitive load after using the resource.

Several reasons may explain such confounding results: (i) the difficulty and specificity of the anatomical topic to be studied; (ii) students' individual differences, notably spatial and motor abilities; and finally, (iii) the interaction between the learner's abilities and the instructional tool (Nguyen et al., 2012; Hoyek et al., 2014). According to the compensating hypothesis, 3D-multimedia resources allow students with low spatial ability to better build their mental model of the anatomical structures (Mayer, 2002; Hegarty & Kriz, 2008). Conversely, the enhancer hypothesis (Hegarty & Sims, 1994; Hegarty, 2005) states that high spatial students are better equipped to process 3D resources as they have enough cognitive capabilities for building efficient mental models. In line with the above-mentioned hypotheses, researchers have been suggesting that learning requires a strong interaction with the environment in general and with digital tools in particular (Nguyen et al. 2012). This is called embodied cognition and it includes all cognitive processes that are linked to the body's interaction with the environment (Jang, 2010).

Embodied cognition is gaining traction and its efficiency in learning has been demonstrated in a wide variety of sciences (e.g., Barsalou, 2008; Cook et al., 2017). This embodied perspective is particularly relevant for learning anatomy as the knowledge to acquire is related to the body itself. Augmented reality (Jain et al., 2017) and virtual reality and immersive anatomy (Weyhe et al., 2018) tools were developed to enrich students' experience resulting in a positive feedback (Hoang et al., 2017). Bauer et al. (2017) developed a mirror-like augmented reality allowing students to move while observing their anatomical structures reaction in real time. Such educational technologies enable researchers and teachers to include more gestures and body movements into learning anatomy. Nevertheless, such initiatives remain scarce and their logistics and developments are expensive.

Triggering embodied cognition during anatomy learning does not necessarily require sophisticated multimedia tools, e.g.: tracing an arrow on a diagram or pointing a word in a text (Macken & Ginns, 2014), facial expressions and hand or forearm movements (Cherdieu et al., 2017; Dickson & Stephens, 2015; Oh et al., 2011), body painting (McMenamin, 2008), drawing on t-shirts (Skinder-Meredith, 2010). All these techniques enhanced learning outcomes. In the STEM (Science Technology Engineering Mathematics) field, Johnson-Glenberg and colleagues (Johnson-Glenberg, 2018; Johnson-Glenberg & Megowan-Romanowicz, 2017) proposed a taxonomy to assess the amount of embodiment in STEM lesson, ergo the amount of sensori-motor engagement, the congruency between gesture and content, and the amount of immersion experienced by the user.

This embodied perspective is particularly relevant for learning musculo-skeletal functional anatomy as the knowledge to be acquired is the movement itself. Movement execution (enacted encoding) during learning and memorizing action sentences results in better performance than simply memorizing the sentences (verbal encoding) (Macedonia & Mueller, 2016). Understanding musculo-skeletal functional anatomy requires not only spatial abilities but also motor imagery which is the ability to imagine a human movement without any real movement execution (Jeannerod, 1994). We used spatial abilities and motor imagery processes to design, develop and validate the efficiency of hundreds of 3D anatomy animations (Hoyek et al., 2014; Berney et al., 2015). However, our 3D animations lacked interaction as they are passively visualized by students. Based on all previous studies and given the place of both movement and interaction in functional anatomy learning we developed a new interactive 3D tool called Antepulsio®. While there are hundreds of applications that provide animations of muscle movement, Antepulsio® is the first application that allows student to analyze and recreate a movement and see feedback on the correctness of the choice. Such manipulation of anatomical structures helps learners to map structures to their own bodies' coordinate systems (Amorim et al., 2006). As stated by Jang and colleagues (2017), "objects that are perceived to be anatomical in nature may prime a more embodied approach to mental imagery than abstract figures." (p. 152). In this perspective, Antepulsio® exercises have a strong potential to trigger embodied cognition. The aim of this paper is to present Antepulsio® instructional design and its research and development approach. Users' experience (UX) tests using the Attrakdiff questionnaire (Hassenzalh, 2003; Lallemand et al., 2015) were conducted. Students' spatial and motor imagery abilities as well as their anatomy examination results were collected. Based on our theoretical and literature review, our research aims to study the relationship between users' experience, anatomy score and spatial and motor imagery abilities.

It is noteworthy to state that this paper does not study the impact of implementing Antepulsio® or to measure embodiment as a mediating variable for students' learning outcomes. Rather, we focus on how improving UX may impact students' performance and how it interacts with some embodied cognition-related variables such as spatial and motor imagery. Hence, the following research questions (RQ) were developed:
RQ1: Does enhancing the application impact the UX?
RQ2: Does the UX have any impact on anatomy scores?
RQ3: Do students' spatial and motor imagery abilities interact with their UX?
RQ4: Do students' spatial and motor imagery abilities interact with their anatomy examination results?
Answering these questions will help teachers and developers in their general Research and Development (R&D) process especially through introducing UX when designing an anatomy 3D tool.

2. General method

2.1. An iterative design process using UX

We used an evidence-based tool design strategy. This process is at the core of our agile manufacturing. Following the development of Antepulsio®, we conducted two experiments. The pilot study aimed at collecting the baseline UX (Baseline 1 – B1) from our first prototype using the Attrakdiff2 (Hassenzalh, 2003; Lallemand et al., 2015) questionnaire (T0). This first version of the tool was evaluated by students enrolled in our functional anatomy course (2018–2019 academic year). Students' official anatomy exam scores were collected before the UX experiment to verify if previous knowledge yielded any effect on UX. Based on UX evaluation and students' feedback, we improved the tool during a redesign process.

Experiment I aimed at testing Antepulsio®'s UX enhancement in a pretest-posttest paradigm. A new student cohort enrolled in our functional anatomy course (2019–2020 academic year) assessed Antepulsio®'s UX (T1: pre-test). They were then enrolled into a one week-long training session using Antepulsio®. The tool's UX was finally tested at (T2 post-test) after the training sessions to verify if any differences with T1 were to be observed. All participants signed an informed-consent form before starting the study. Our local management and ethics committee approved the experimental design after the experimenters presented the objectives and procedures to the scientific board council.

2.2. Anatomical and pedagogical content of Antepulsio® application

The game engine Unity® was used to develop Antepulsio®. A real-time physical simulation model of the human body allowing realistic motions was used to replicate muscle contraction. The three types of muscular contraction could be reproduced by the model: eccentric, static and concentric. The following presentation of the tool is a general description of the four different types of exercises that were developed. More information on the differences between the first and second version of the tool and design decisions is available in section 3.2 and 6.2.

2.2.1. Exercise 1: Muscle understanding

The aim of this exercise is to associate a muscle to a movement that it produces. The learner has to identify a muscle on a static position of the 3D model and then to visualize the different muscular insertion points. This allows for displaying the details of the muscle localization and or users to predict its movement following contraction. Lastly, the student can observe the muscle action in motion.

2.2.2. Exercise 2: Movement analysis

The learner must observe a movement and its kinematics and then execute the movement. The learner has to divide the movement into several components by using the appropriate movement terminology and locating the mobile against the immobile body segment. In the second stage the learner has to analyze the type of muscle contraction needed for that movement. Finally, the learner needs to identify the agonist responsible muscles.

2.2.3. Exercise 3: Movement reproduction

The knowledge and skills outlined in the two previous exercises (i.e., kinematics, type of contraction, muscles function, terminology) need to be acquired to complete this exercise. After watching a movement, the learners have first to execute the movement and then to reproduce it precisely in the application. After choosing all the appropriate muscles with their respective contraction type the student can launch the 3D simulation in order to verify whether the movement they reproduced is identical to the original one (see Figure 1).

2.2.4. Exercise 4: Assessment

This summarizing exercise is a multiple-choice questionnaire. It allows the learners to evaluate themselves and get feedback on their comprehension level.

Figure 1. Validation step of the exercise 3: Movement reproduction. The student here made a mistake by choosing the wrong muscles and wrong contraction type while analyzing slow trunk flexion in standing position

3. Pilot study

3.1. Sample and procedure

Forty-five students (7 females; age: 17.89 ± 0.83 years) enrolled in our first-year kinesiology program at Lyon 1 University (2018– 2019 academic year) participated on a voluntary basis in this first experience. After following an entire functional anatomy course (fall 2018) they tested the first version of Antepulsio® in April 2019 (B1). For more information on the curriculum design and the 3D tools used during the semester see Hoyek et al. (2014). None of them had previously tested the Antepulsio® tool. They tested the app individually in a quiet room. Students tested once all four types of exercises and their corresponding steps (see section 2.2). To pass to the following step/exercise they had to validate the previous one. They had no time constraint. They had an unlimited number of tries. An average of 20 minutes was needed to complete the exercises. Students' feedback on the tool's properties (ergonomics, usability, ease of use, etc.) was collected during focus groups after the completion of each exercise and at the end of the session. Finally, the Attrakdiff2 (Hassenzalh, 2003; Lallemand et al. 2015), UX measurement tool was administered.

3.2. Material

3.2.1. Antepulsio® version 1

The first version consisted of the above-mentioned exercises. The learning path was linear and progressive (see Figure 2). Students could not pass to the following exercise before validating the previous one. Students were guided in every single step through written instructions. Furthermore, the User Interface contained the following features: (i) buttons and textual information were positioned in several locations on the screen; (ii) different font types and sizes were used; (iii) the remaining steps of an exercise were permanently present on the screen.

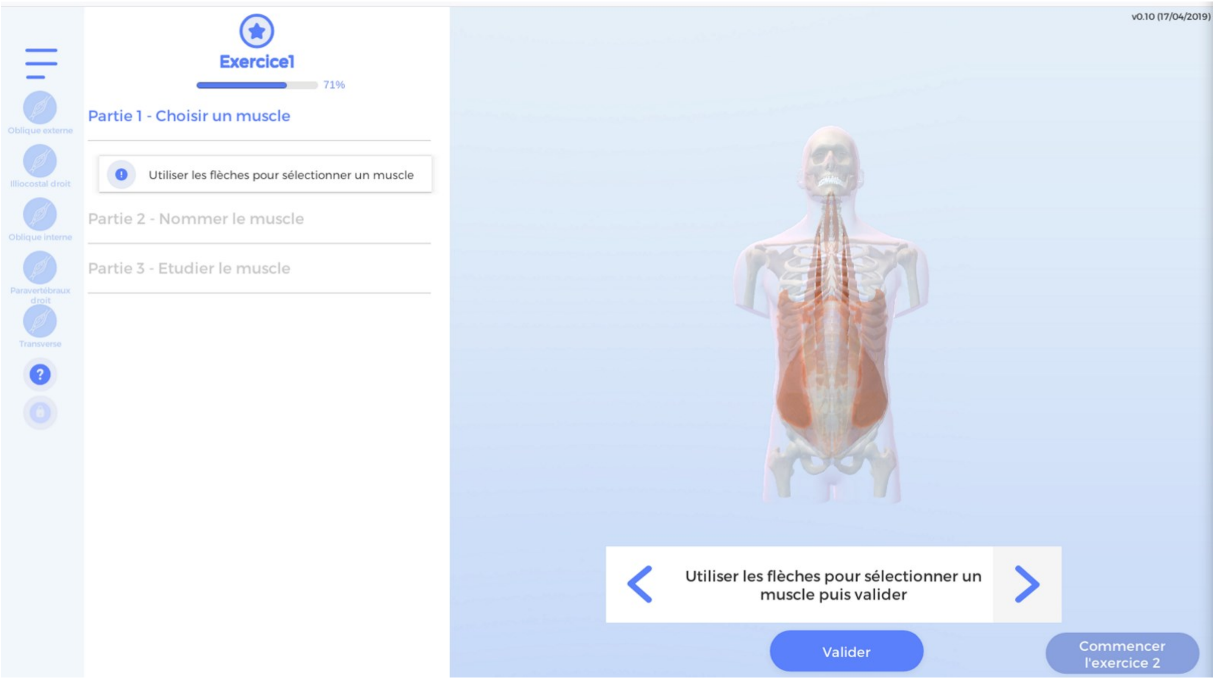


Figure 2. Antepulsio® version 1-Exercise 1: Muscle understanding

3.2.2. Students' feedback

Oral feedback from the students was collected regarding each of the application exercises. The scientific assistant noted down the feedback given by each subject in a table. At the end of the session they had to answer the three following questions: (i) What is your general opinion regarding Antepulsio®; (ii) Would you use Antepulsio® for studying anatomy? (iii) What could be enhanced in the application? The answers/feedback for each question are used to support our quantitative findings.

3.2.3. UX

Attrakdiff2 (Hassenzalh, 2003; Lallemand et al. 2015), a quantitative UX measurement tool, evaluates the hedonic and pragmatic qualities of an interactive system. It is a standardized questionnaire with four subscales of seven items each, for a total of 28 items. The items are in the form of pairs of contrasting words to be assessed using 7-point Likert scales, ranging from -3 to +3. The order in which the items are administered is standardized and the items are mixed. The Attrakdiff subscales are as follows : Pragmatic Quality (PQ) describes the usability, the usefulness of the product and indicates how well the product enables users to achieve their goal in completing task ; Hedonic Quality-Stimulation (HQ-S) indicates the extent to which the product can support the need for stimulation ; Hedonic Quality -Identification (HQ-I) indicates the extent to which the product allows the user to identify with it; and the Appeal (APP) , or the global attractiveness is a value of the product based on the perception of pragmatic and hedonic qualities. We measured the average of each subscale for each completed questionnaire by the students.

3.2.4. Anatomy score

We collected the scores obtained by each participant during their official anatomy exam (fall 2018 session) which happened before the student's use of the tool and their UX measurements. It is a standardized 20-item questionnaire with three possible answers: true, false and "don't know". We recorded the correct answers.

4. Results

4.1. UX properties

Attrakdiff constructs Cronbach's alphas ranges from .60 to .83. They yielded normal distributions (.086 < $D_{K\text{-}S}$ < .200; p > .05). Anatomy Score (AS) also yielded a normal distribution: M = 11.59; SD = 3.98; $D_{K\text{-}S}$ = .110; p > .05. Confirmatory Factorial Analyses to verify models fitness with our data couldn't be ran as the ratio between N and the model's number of parameters (9 in our case) should be from 5 to 10 cases for each parameter (Bentler & Chou, 1987; Bollen, 1989; cited by Kyriazos, 2018). As our ratios are 5 for B1 and 1.89 for T1 and T2, CFAs were not calculated. Furthermore, it is generally preferable to conduct CFAs with large samples (Brown, 2015; Kline, 2016; cited by Kyriazos, 2018). As in Hassenzahl and colleagues' original tool validation study (2013) and in Lallemand and colleagues' French validation study (2015), we found that both hedonic scales are correlated (r = .422; p < .05), PQ is also correlated with HQ-S (r = .468; p < .05) and HQ-I (r = .439; p < .05) hedonic scales. On the other hand, a regression analysis confirmed that PQ, HQ-S and HQ-I positively predict APP: β = .465; p < .001; R^2 = .607. Descriptive results and Kolmogorov-Smirnov normality tests ate shown in Table 1.

Table 1. Attrakdiff dimensions and Anatomy score descriptive results and Kolmogorov-Smirnov normality tests at Baseline 1 (B1)

	B1 (N = 45)			
	M	SD	$D_{K\text{-}S}$	α
PQ	1.02	.61	.086	.60
HQ-S	1.63	.79	.092	.78
HQ-I	.78	.74	.095	.72
UX	1.14	.54	.097	.80
APP	1.82	.69	.123	.83
Anatomy Score	11.59	3.98	.110	

UX subdimensions ranged from M = .78; SD = .74 (for HQ-I), to M = 1.63; SD = .79 (for HQ-S). The overall UX means is: M = 1.14; SD = .54 and the tool's overall appeal is M = 1.82; SD = .69. All values are above 0, which indicate a satisfying tool's UX assessment from the students.

4.2. UX and anatomy score

All correlation tests between the Attrakdiff dimensions (including global UX score) and Anatomy Score display non-significant results (p > .05). We further investigated for a potential link between anatomy scores and UX by breaking the sample down into a low-level group and a high-level group using the anatomy score median as a cutoff point (Md = 11). One-way ANOVA F-tests between group membership and all Attrakdiff dimensions yielded no significant differences (p > .05).

5. Discussion

Our first results provided evidence that all Attrakdiff subscales are intercorrelated; and pragmatic and hedonic qualities contribute to the tool's global appeal. These results are in line with the theoretical model description of Attrakdiff (Hassenzalh, 2003; Lallemand et al., 2015). Using the Attrakdiff2 is thus appropriate in our R&D process. Furthermore, our students found this first version attractive. It is noteworthy to point out that students had already attended a semester during which they learned anatomy using our classical 3D animations (see Hoyek et al., 2014). Giving the efficiency of our existing 3D animations (Hoyek et al., 2014), this gives more value to the global attractiveness of Antepulsio®.

Furthermore, students' verbal feedback was considered in the development of the second version of the application. The main positive feedback was: the interaction with the avatar enables a better 3D visualization of movement facilitating motor imagery; the interaction makes the students more active in their learning process; a good tool for revision and for assessments. The main negative feedback was: too much guidance in order to pass from an exercise to another; some guidelines were not clear; some exercises were lacking immediate feedback on the correctness of the response; legends explaining the colors associated to muscle contraction type were missing; exercises required a lot of time to be completed; small software bugs (see section 6.2).

Finally, the pilot study does not reveal any correlation between the Attrakdiff dimensions and students' anatomy scores. Using a median split, lower and higher knowledge student rated Antepulsio® equally. More complex interaction between anatomy scores and the enhanced version of Antepulsio® is discussed in Experiment I.

6. Experiment I

6.1. Sample and procedure

A year later, a second improved version of Antepulsio® was tested by a new cohort of 17 students (6 females; 18.07 ± 0.80 years) in October 2019. Students were enrolled into our functional anatomy course (2019–2020 academic year). They were also exposed to our previous 3D animations and were enrolled into similar curriculum design (see Hoyek et al. 2014). The experimental protocol consisted of a pre and post-training paradigm implemented into a semester. Students participated in three training sessions using Antepulsio®. They tested individually in a quiet room the application. At the beginning of session 1, students completed spatial ability and motor imagery tests. Session 1 consisted of 7 thematic modules (see 6.2 section below) and lasted 25 minutes and was programmed to stop automatically when time ran out. At the end of Session 1, the Attrakdiff2 questionnaire was administered (T1). Session 2 contained 8 different thematic modules and lasted 20'. Session 3 contained 7 new different thematic modules and lasted 20'. At the end of Session 3, the Attrakdiff2 questionnaire was administered (T2). Students had up to 3 tries/step to answer each question. One night, at least, separated each session.

6.2. Material

6.2.1. Antepulsio® version 2

Version 2 consisted of the same above-mentioned exercises (see Figure 3). However several improvements were made. The learning path was reorganized into thematic modules (e.g., concentric muscle contraction; abdominals; ball throwing analysis; etc.). However those latter were not always organized from Exercise 1 to 4 in a linear way. Students could navigate by choice, that is they could postpone an exercise, pass to another one and come back later to complete it. A timer was inserted in the upper right corner. Several modifications were made to the User Interface: (i) buttons and textual information were positioned in the left part of the screen; (ii) font types, sizes and colors were better organized and unified across exercises; (iii) only the title of the exercise's step was shown along with a percentage of progression giving students more information on the remaining steps to complete the exercise.

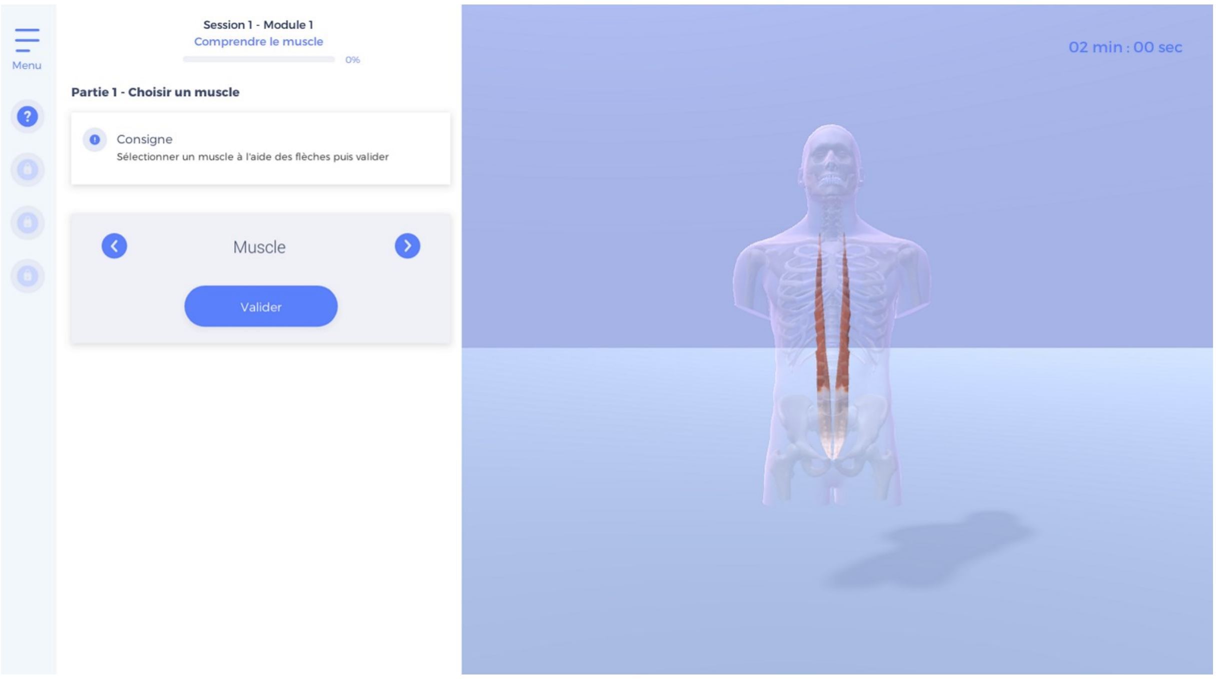


Figure 3. Antepulsio® version 2-Exercise 1: Muscle understanding

6.2.2. UX

The Attrakdiff2 questionnaire was administered to compare scores to the original Baseline scores from version 1.

6.2.3. Spatial ability

Spatial Ability (SA) was measured using the Vandenberg and Kuse Mental Rotation Test (MRT: Vandenberg and Kuse, 1978). It is a paper-and-pencil test that consists of 24 items of 3D-objects. Each item consists of 5 figures, a 3D model on the left and 4 on the right among which participants must indicate those that are similar to the model. Participants have to mentally rotate the target figures to find the two correct items that match the reference. The score can range from 0 to 24. One point is given for an item only if both correct test figures were identified. The test was to be completed within a 6-minute period. This test was performed before training.

6.2.4. Motor imagery

Motor Imagery (MI) was measured using a laterality judgement test. It consists of a computer-based test that presents different hand stimuli on the screen: right or left hand, back or palm hand in various orientations (0°, 90°, 180°, 270°). Participants had to find the laterality of a stimulus without any hand or head movements by pressing on the keyboard's left or right arrow. The number of correct answers was collected. This test was carried out before training.

6.2.5. Anatomy Score (AS)

Like the pilot study, the correct answers of 20-items that were randomly selected from the same database were recorded. The anatomy test was run a second time at T2. The goal was to verify for any significant differences in anatomy knowledge after the one-week training using the tool, and if the knowledge change interacted significantly with UX.

7. Results

7.1. Descriptive results

Results at T1 and T2 indicate higher UX means than at B1 (Table 2). The normality assumptions for parametric tests were met for all constructs ($.127 < D_{K\text{-}S} < .897$; $p > .05$), except for MI ($D_{K\text{-}S} = .674$; $p < .05$). Moreover, UX means at T2 are higher than at T1. T-tests will be conducted to verify if these differences are significant.

Table 2. Attrakdiff dimensions, mental imagery abilities, Anatomy score results and Kolmogorov-Smirnov normality tests at T1 and T2

	T1 (N = 17)			T2 (N = 17)		
	M	*SD*	$D_{K\text{-}S}$	*M*	*SD*	$D_{K\text{-}S}$
PQ	1.32	.61	.158	1.39	.62	.154
HQ-S	1.73	.66	.184	1.85	.70	.225
HQ-I	1.64	.68	.127	1.85	.74	.182
UX	1.50	.38	.209	1.62	.45	.157
APP	1.79	.68	.216	1.94	.46	.198
Spatial Ability	5.47	2.67	.924			
Mental Imagery	61.47	3.67	$.674^{*}$			
Anatomy Score	12.47	2.76	.897	14.41	4.88	.935

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

7.2. UX scores change between B1 and T1

We conducted independent *t*-tests between T1 and B1 for Attrakdiff evaluations (i.e., the delta - Δ) to verify significant differences, as samples are independent (see Table 3). Results display significant deltas between B1 and T1 for HQ-I only with a large effect size ($t = 4.17$; $p < .001$; $d = 1.18$); as well as for UX ($t = 2.51$; $p = .015$;

d = .71) for a medium to large effect size. Even if PQ change's *p* value is non-significant, it still yields a medium effect size. Attrakdiff dimensions' change yield an overall significant and large UX change (*Δ* = .36).

Table 3. Independent *t*-tests between T1 and B1

	Δ	*t*	*p*	*d*
PQ	.30	1.72	.090	.49
HQ-S	.10	.46	.645	.13
HQ-I	.86	4.17	.001	1.18
UX	.36	2.51	.015	.71
APP	-.03	-.15	.879	-.04

7.3. UX scores change between B1 and T2

Independent *t*-tests were also conducted between B1 and T2, for the same reasons as mentioned above (table 4). In contrast with the T1-B1 comparison, all B1-T1 Attrakdiff dimensions deltas display significant changes except for HQ-S. In order of descending effect size: HQ-I (*Δ* = 1.07; *t* = 5.07; *p* < .001; *d* = 1.44); APP (*Δ* = .14; *t* = 3.88; *p* < .001; *d* = 1.10); PQ (*Δ* = .37; *t* = 2.12; *p* = .038; *d* = .60) and HQ-S (*Δ* = .22; *t* = 1; *p* = .318; *d* = .28). These results lead to a significant overall UX score change: *Δ* = .48; *t* = 3.25; *p* = .002; *d* = .92; with a large effect size.

Table 4. Independent *t*-tests between T2 and B1

	Δ	*t*	*p*	*d*
PQ	.37	2.12	.038	.60
HQ-S	.22	1.00	.318	.28
HQ-I	1.07	5.07	.001	1.44
UX	.48	3.25	.002	.92
APP	.14	3.88	.001	1.10

7.4. UX scores change between T1 and T2

We ran paired-samples *t*-tests between T2 and T1 for our test-posttest sample (table 5). The results indicate no significant change between T1 and T2 for any of the Attrakdiff dimensions, nor for the overall UX dimension (all p-values are above the .05 confidence interval). Despite a non-significant *p* value, we can still observe – small – effect sizes for HQ-I and UX's deltas (respectively: *Δ*M = .20; *d* = .23 and *Δ*M = .11: *d* = .17).

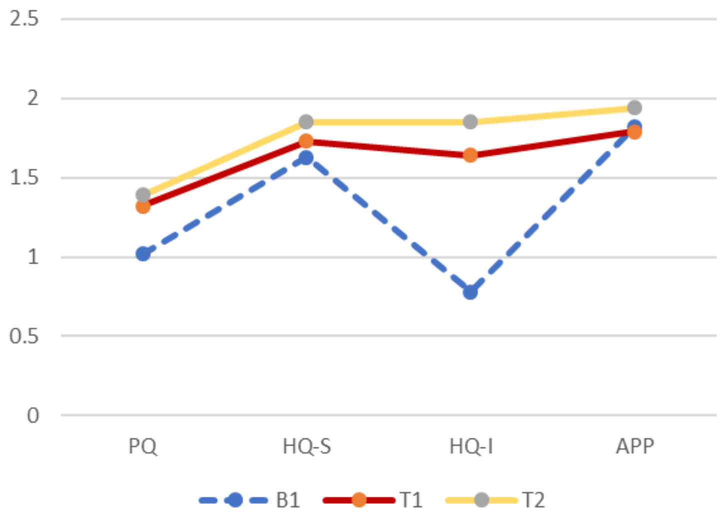


Figure 4. UX assessments at B1, T1 and T2

Table 5. Paired *t*-tests between T2 and T1

	ΔM	*t*	*p*	*d*
PQ	.07	.55	.589	.09
HQ-S	.11	.55	.587	.13
HQ-I	.20	1.44	.168	.23
UX	.11	1.24	.230	.17
APP	.14	1.00	.332	.19

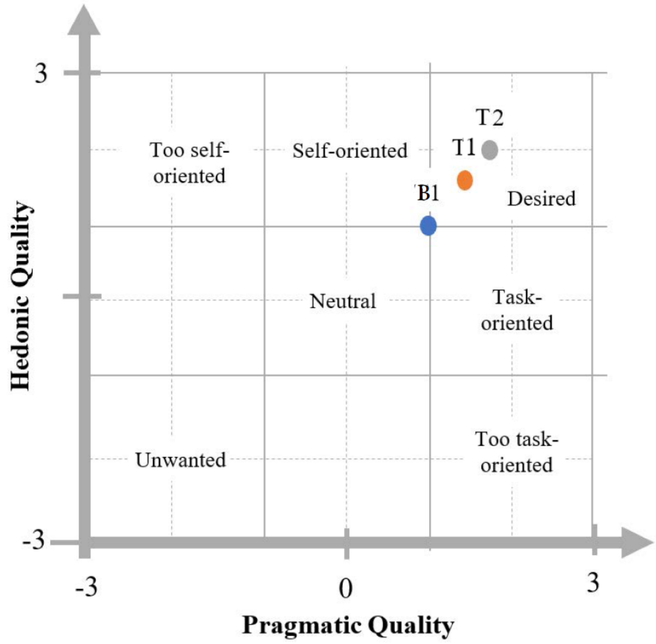


Figure 5. UX change between B1 and T23

Using the UX framework based on Attrakdiff, as proposed by Lallemand et al. (2015) and inspired by Hassenzahl (2003, p. 37), Figures 4 and 5 show a qualitative enhancement of Antepulsio®'s UX during our experimental protocol. The most significant enhancement was between pilot study and experiment 1, especially for the HQ-I dimension. At B1, UX was positioned at the intersection between four zones: The Neutral, the Self-Oriented, the Task-Oriented and the Desired zones. Then, at T1 and T2, Antepulsio®'s UX assessments yielded a complete entry into the Desired zone, which indicates a balance between the tool's hedonic qualities and pragmatic qualities. In other words, the system provides strong enjoyment qualities, as well as strong effectiveness.

7.5. Anatomy score change between T1 and T2

The anatomy score's delta between T1 and T2 (Δ_{AS}) score yielding a normal distribution ($D_{K\text{-}S}$ = .139; *p* = .200), we ran a paired-sample *t*-test with our test-posttest population to verify if the one-week training significantly improved the anatomy score (AS) . The result indicates a significant and positive change: Δ_{AS}M = 1.94; Δ_{AS}SD = 3.03; *t* = 2.64; *p* = .018; *d* = .77.

7.6. Anatomy score and UX

To run tests on Anatomy Score (AS), we used Δ_{AS} to verify if the training-post-training AS difference significantly interacted with any of the Attrakdiff dimensions. We found that the general Hedonic Quality mean (HQ) at T2 is negatively correlated with Δ_{AS}: *r* = -.500; *p* = .042. Whereas the pragmatic perception (PQ) yields no interaction with performance.

7.7. Motor imagery, spatial ability, UX and performance

Embodiment-linked processes were also verified. Mental Imagery at T1 was found to predict PQ at T2 (i.e. by the end of the experiment), explaining 20% of PQ's variance: $\beta = .491$; $p < .05$; $R^2 = .200$. Consequently, MI was also found to predict the overall UX score at T2, explaining 23.50% of UX's variance: $\beta = .532$; $p < .05$; $R^2 = .235$

On the other hand, Spatial Ability was found to predict Δ_{AS}, explaining 33% of Δ_{AS}'s variance: $\beta = .575$; $p < .05$; $R^2 = .330$. Consequently, we ran regression analyses with Δ_{AS} as a predicted variable, to investigate predictive models. We executed Model 1, which comprises only PQ and HQ as UX dimensions predictors at T2. We also executed Model 2, with PQ and HQ, to which we added SA, as predictors. Results of regression models with ΔAS as a predicted variable are shown in Table 6.

Table 6. Regression models with Δ_{AS} as a predicted variable at T2

	Model 1		Model 2	
Predictors	β	p	β	p
Constant	6.08	.035	7.83	.005
PQ	.220	.356	.236	.253
HQ	-.547	.033	-.432	.049
Spatial Ability			.491	.028
F	2.90	.088	4.68	.012
R^2	19.20%		41%	

In Model 1, HQ at T2 significantly predicts Δ_{AS}: $\beta = -.547$; $p = .033$. Whereas PQ (T2) do not yield a significant effect within the model ($p = .356$). Total explained variance for Model 1 is $R^2 = 19.20\%$. In Model 2, HQ at T2 significantly predicts Δ_{AS}: $\beta = -.432$; $p = .049$. SA improves the model's variance by a $\Delta R^2 = 21.80\%$, and significantly predicts Δ_{AS}: $\beta =.491$; $p = .028$. Consequently, Model 2 explains $R^2 = 41\%$ of Δ_{AS}' variance. Hence, results indicate that both HQ and SA predict the change of the anatomy score between T1 and T2.

8. Discussion

The test-posttest design of Experiment I provided a more complete overview of Antepulsio®'s UX. First, our results provided evidence of a significant and large UX improvement especially for the HQ-I dimension between the second version of the application (T1 & T2) and the first prototype (B1). Thus, the enhancement made following Experiment I had a positive impact on students' experience (UX) and appreciation (RQ1). The increase of HQ-I may be explained by the guidelines enhancements and the additional legends associated with colors explaining the type of muscle contraction. This increase might also be explained by the fact that HQ-I had the lowest score at B1 ($M = .78$) giving a large space for improvement. These enhancements may have led students to better identify themselves to the application in general and to the 3D avatar in particular. We assume that the legends and visual cues have helped students in their 3D visualization of anatomical structures. This is in line with Roach et al. (2018) study who provided evidence that guiding students where to look improves their spatial reasoning. The authors suggested that visual guidance may be applied in anatomy to improve student's interpretation of visual content such as anatomical structures.

We did not find any significant change for any of the Attrakdiff dimensions between T2 and T1. This result is mainly explained by the fact that the same application was tested two times by the same students within the same week. Furthermore, this result suggests that there is no effect of three training sessions on UX. In other words, learning about the content conveyed by the application and learning how to use the application itself did not significantly improve Antepulsio®'s UX scores. This result is at odds with previous studies (Bhattacherjee et al., 2004; Venkatesh et al., 2011; Martin et al., 2016) where interaction with an application significantly decreases UX after use. The fact that Antepulsio®'s UX did not decrease after use gives more value of its acceptance by students.

The fact that the HQ-I dimension has been found to evolve in comparison with the other Attrakdiff dimensions says something about the evoked cognitive processes when using Antepulsio®. We assume that interactions with a human avatar enables user's identification. Several cognitive sub-processes may have been called upon in this very complex self-identification. For instance, individuals express their self through physical objects (Prentice, 1987; Hassenzahl, 2003). We assume that interacting with a human avatar calls upon empathy (Hamilton-Giachritsis et al., 2018). Mental and motor imagery, action observation and embodiment are part of empathy components (Decety & Jackson 2004). One sees others through one's own embodied cognition (Decety and Jackson 2004). When using Antepulsio®, students are turning the avatar in 3D, imagining the movement to study and observing the avatar's movement, making several clicks to interact with the avatar. All these processes share similar underlying neurocognitive processes (Decety & Jackson 2004; Vogt et al., 2013) and may be behind this significant increase of HQ-I.

Unlike the pilot study, we found an interaction between students' UX and their anatomy scores (RQ2). More precisely, the hedonic quality mean at T2 was negatively correlated with the anatomy scores. This result indicates that anatomy scores interact with a decrease of the tool's hedonic perception, whereas the pragmatic perception yields no interaction with performance. Learning Antepulsio®'s conveyed content decreased the general pleasure using the tool. We can speculate that when the learning goal is reached, the hedonic perceived quality wears out. In other words, the instrumental interest and the enjoyment decreases, whereas the tool's pragmatic qualities are equally perceived as it helped participants reach their learning goals. Indeed, as stated by Hassenzahl (2003, p. 34), "a pragmatic product is primarily instrumental. It is used to fulfil externally given or internally generated behavioral goals." Hedonic quality, however, focuses on "the Self, i.e., the question of why does someone (...) use a particular product" (Hassenzahl, 2010, p. 50). This is in line with two of students feedbacks. First, they stated that Antepulsio® is a good tool for revision. This gives us clear indication regarding its future implementation in our curricular design in order to keep the hedonic stimulation as high as possible. Secondly, students stated that some exercises required a lot of time to be completed. Even though we shortened the exercise length, perhaps more of the exercises could be shorter and more pleasant.

Finally, motor imagery predicts PQ and overall UX score at T2 (RQ3). To our knowledge, this is the first time that motor imagery ability is linked to UX in general and to its pragmatic dimension. It is thus difficult to discuss such a novel finding with regards to previous studies. We can however speculate that good motor imagery ability is needed to positively estimate the pragmatic dimension of Antepulsio®. Therefore, our learning tool may be more adequate for high motor imagery students and confirms thus the enhancer hypothesis (Hegarty and Sims, 1994; Hegarty, 2005). This interpretation remains speculative and more research is needed before drawing final conclusions. On the other hand, mental rotation ability predicts anatomy score improvement (RQ4). These results are in line with a huge amount of previous studies ascertaining the importance of good spatial abilities in anatomy knowledge acquisition (e.g. Guillot et al., 2007).

9. Conclusion and implications for design

Fostering embodied learning in functional anatomy is simple because the main knowledge to acquire is the movement itself and its analysis. It can occur through real movement execution, motor imagery or action observation. These latter share similar processes and similar neural substrates (Vogt et al., 2013).

An additional way to foster embodied learning in anatomy passes through real movement execution (exercises 2 and 3) and interactions with the avatar. The clicks and manipulations of the avatar, the perspective changes, the zoom in and out actions are made to better visualize the human body and its movements. According to Wiedenbauer and Jansen-Osmann (2008) manual rotation training enhances mental rotation performance. We thus assume that giving the students the ability to interact with the avatar would enhance their spatial ability and reduce their cognitive load.

Nevertheless, even with such interactions, our results provided evidence that good motor imagery ability is needed to positively estimate the pragmatic dimension of Antepulsio®. The conception of an anatomy digital tool should consider such individual differences as well as students' cognitive load and visual strategies (Mayer, 2002; Hegarty & Kriz, 2008; Hegarty & Sims, 1994; Hegarty, 2005). This would help developers and educators choose the best visual cues to add into their applications to foster empathy and consequently embodied cognition during the learning process.

Another notable result of our study is the importance of students UX during tool development. Our user-centered design method allowed us to improve the app's interface and the exercises gameplay. The enhancement of the Attrakdiff2 scores across our study was noteworthy. Studying in detail the change of its dimensions gave us insight into some of our students' cognitive and psychological learning processes. The overall hedonic dimension can give us insight into the enjoyment perceived by the students. Our students' enjoyment decreased as soon as the pedagogical goals were attained. This will help us enhance our gameplay in the future for a better learning motivation. It will also help us better implement our tool into the curriculum. The hedonic identification dimension may be linked to students' empathy with the 3D avatar and could therefore give insight into the embodied learning process.

Finally, in the absence of a control group, the impact of Antepulsio® on learning outcomes was not studied. The implementation of the tool within our curriculum followed by a randomized controlled study will thus be conducted in the future. Furthermore a quantification of embodiment will be conducted. This will help us better understand the impact of Antepulsio® on complex anatomy concepts understanding.

Acknowledgment

We thank brûle. company who coordinated the production process by interlinking the project's different work packages. We thank the Anatoscope company who built a real-time physical simulation model of the human body. We thank Ochelys company who ensured the agile manufacturing of the application thanks to the UX tests analysis. The R&D Antepulsio® project is funded by European Regional Development Fund (ERDF) – Auvergne-Rhône-Alpes in the context of the FUI 22 (Fonds Uniques Interministériels) call for projects.

References

Abid, B., Douard, R., Hentati, N., Ghorbel, A., Delmas, V., & Chevallier, J.-M. (2007). L'enseignement de l'anatomie entre la craie et la pédagogie numérique. Résultats d'une étude comparative [Teaching Anatomy from chalk to digital pedagogy. Results from acomparative study]. *Morphologie, 91*, 107–108.

Amorim, M. A., Isableu, B., & Jarraya, M. (2006). Embodied spatial transformations: "Body analogy" for the mental rotation of objects. *Journal of Experimental Psychology: General*, *135*(3), 327. doi:10.1037/0096-3445.135.3.327

Barsalou, L. W. (2008). Grounded cognition. A*nnual Review of Psychology, 59*, 617–645. doi:10.1146/annurev.psych.59.103006.093639

Bauer, A., Neog, D. R., Dicko, A.-H., Pai, D. K., Faure, F., Palombi, O., & Troccaz, J., (2017). Anatomical augmented reality with 3D commodity tracking and image-space alignment. *Computers & Graphics, 69*, 140–153. doi:10.1016/j.cag.2017.10.008

Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 16*, 78-117. doi:10.1177/0049124187016001004

Berney, S., Bétrancourt, M., Molinari, G., & Hoyek. N. (2015). How spatial abilities and dynamic visualizations interplay when learning functional anatomy with 3D anatomical models. *Anatomical Sciences Education*, *8*(5), 452-462. doi:10.1002/ase.1524

Bhattacherjee, A., & Premkumar, G. (2004). Understanding changes in belief and attitude toward information technology usage: A Theoretical model and longitudinal test. *MIS Quarterly., 28*, 229–254.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Jon Wiley & Sons. doi:10.1002/9781118619179

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

Cherdieu, M., Palombi, O., Gerber, S., Troccaz, J., & Rochet-Capellan, A., (2017). Make gestures to learn: Reproducing gestures improves the learning of anatomical knowledge more than just seeing gestures. *Frontiers in Psychology, 8*,1689. doi:10.3389/fpsyg.2017.01689

Clunie, L., Morris, N. P., Joynes, V. C. T., & Pickering J. D. (2018). How comprehensive are research studies investigating the efficacy of technology-enhanced learning resources in anatomy education? A Systematic review. *Anatomical Sciences Education., 11*, 303-319.

Cook, S.W., Friedman, H.S., Duggan, K.A., Cui, J., & Popescu, V., (2017). Hand gesture and mathematics learning: lessons from an avatar. *Cognitive Science, 41*, 518–535. doi:10.1111/cogs.12344

Decety, J., & Jackson, P. L. (2004). The Functional architecture of human empathy. *Behavioral and cognitive neuroscience reviews, 3*(2), 71-100.

Dickson, K.A., Stephens, B.W., (2015). It's all in the mime: Actions speak louder than words when teaching the cranial nerves. *Anatomical Sciences Education, 8*, 584–592. doi:10.1002/ase.1531

Guillot, A., Champely, S., Batier, C., Thiriet, P., & Collet, C. (2007). Relationship between spatial abilities, mental rotation and functional anatomy learning. *Advances in Health Sciences Education*, 12, 491–507.

Hamilton-Giachritsis, C., Banakou, D., Quiroga, M. G., Giachritsis, C., & Slater, M. (2018). Reducing risk and improving maternal perspective-taking and empathy using virtual embodiment. *Scientific reports*, 8(1), 1-10.

Hassenzahl, M. (2003). The thing and I: Understanding the relationship between user and product. In M. A. Blyth, A. F. Monk, K. Overbeeke, & P.C. Wright (Eds.), Human-computer interaction series (3). Funology: From usability to enjoyment (pp. 31-42). Dordrecht, Netherland: Springer

Hassenzahl, M. (2010). Experience design: Technology for all the right reasons. *Synthesis lectures on human-centered informatics*, *3*(1), 1-95. doi:10.2200/S00261ED1V01Y201003HCI008

Hegarty, M. (2005). Multimedia learning about physical systems. In R. E. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning* (1st ed., pp. 447–465). Cambridge, UK: Cambridge University Press.

Hegarty, M., & Kriz, S. (2008). Effect of knowledge and spatial ability on learning from animation. In R. K. Lowe, & W. Schnotz (Eds.), *Learning with Animation: Research Implications for Design* (1st ed., pp. 3–29.). New York, NY: Cambridge University Press.

Hegarty, M., & Sims, V. K. (1994). Individual differences in mental animation during mechanical reasoning. *Memory & Cognition, 22*, 411–430.

Hoang, T., Reinoso, M., Joukhadar, Z., Vetere, F., Kelly, D., (2017). Augmented Studio: Projection Mapping on Moving Body for Physiotherapy Education. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17*. Association for Computing Machinery, Denver, Colorado, USA, pp. 1419–1430. https://doi.org/10.1145/3025453.3025860

Hopkins, R., Regehr, G. & Wilson, T.D. (2011). Exploring the changing learning environment of the gross anatomy lab. *Academic Medicine, 86*, 883–888.

Hoyek, N., Collet, C., Di Rienzo, F., De Almeida, M. & Guillot, A. (2014). Effectiveness of three-dimensional digital animation in teaching human anatomy in an authentic classroom context. *Anatomical Sciences Education*, *7*(6), 430-437. doi:10.1002/ase.1446

Jain, N., Youngblood, P., Hasel, M., & Srivastava, S. (2017). An Augmented reality tool for learning spatial anatomy on mobile devices. *Clinical Anatomy, 30*, 736–741. doi:10.1002/ca.22943

Jang, S., (2010). *From Vesalius to virtual reality: How embodied cognition facilitates the visualization of anatomy.* New York, NY: Columbia University.

Jang, S., Vitale, J. M., Jyung, R. W., & Black, J. B. (2017). Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment. *Computers & Education*, *106*, 150-165. doi:10.1016/j.compedu.2016.12.009

Jeannerod, M. (1994). The Representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences, 17*, 187–202.

Johnson-Glenberg, M. C., & Megowan-Romanowicz, M. C. (2017). Embodied science and mixed reality: How gesture and motion capture affect physics education. *Cognitive Research: Principles and Implications, 2*(24). doi:10.1186/s41235-017-0060-9

Johnson-Glenberg, M. C. (2018). Immersive VR and education: Embodied design principles that include gesture and hand controls. *Frontiers in Robotics and AI, 5*(81). doi:10.3389/frobt.2018.00081

Khalil, M. K., Johnson, T. E., & Lamar, C. H. (2005). Comparison of computer-based and paper-based imagery strategies in learning anatomy. *Clinical Anatomy, 18*, 457– 464.

Lallemand, C., Koenig, V., Gronier, G., Martin, R. (2015). Création et validation d'une version française du questionnaire AttrakDiff pour l'évaluation de l'expérience utilisateur des systèmes interactifs [Creating and validating a french version of the AttrakDiff questionnaire for evaluating user's experience of interactive systems]. *Revue européenne de psychologie appliquée 65*, 239-252.

Lewis, T. L., Burnett, B., Tunstall, R. G., & Abrahams, P. H. (2014). Complementing anatomy education using three-dimensional anatomy mobile software applications on tablet computers. *Clinical Anatomy, 27*, 313–20.

Lim, K. H., Loo, Z. Y., Goldie, S. J., Adams, J. W., & McMenamin, P. G. (2016). Use of 3D printed models in medical Education: A Randomized control trial comparing 3D prints versus cadaveric materials for learning external cardiac anatomy. *Anatomical Sciences Education, 9*, 213–21.

Macedonia, M., & Mueller, K. (2016). Exploring the neural representation of novel words learned through enactment in a word recognition task. *Frontiers in Psychology*, *7*, 953. doi:10.3389/fpsyg.2016.00953

Macken, L., & Ginns, P., (2014). Pointing and tracing gestures may enhance anatomy and physiology learning. *Medical Teacher, 36*, 596–601. doi:10.3109/0142159X.2014.899684

Martin N., Jamet É., Erhel S., & Rouxel G. (2016). From acceptability to acceptance: Does experience with the product influence user initial representations? In C. Stephanidis (Eds.), *International Conference on Human-Computer Interaction* (pp. 128-133). doi:10.1007/978-3-319-40548-3_21

Mayer, R. E. (2002). Cognitive theory and the design of multimedia instruction: An Example of the two-way street between cognition and instruction. *New Directions for Teaching and Learning, 2002*, 55–71.

McMenamin, P. G. (2008). Body painting as a tool in clinical anatomy teaching. *Anatomical Sciences Education, 1*, 139–144. doi:10.1002/ase.32

McMenamin, P. G., Quayle, M. R., McHenry, C. R., & Adams, J. W. (2014). The Production of Anatomical Teaching Resources Using Three-Dimensional (3D) Printing Technology. *Anatomical Sciences Education, 7*, 479–86.

Mitsuhashi, N., Fujieda, K., Tamura, T., Kawamoto, S., Takagi, T., & Okubo, K. (2009). BodyParts3D: 3D structure database for anatomical concepts. *Nucleic acids res, 37*(suppl 1), D782-D785.

Moro, C., Štromberga, Z., Raikos, A., & Stirling, A. (2017). The Effectiveness of virtual and augmented reality in health sciences and medical anatomy. *Anatomical Sciences Education, 10*, 549–59.

Nicholson, D. T., Chalk, C., Funnell, W. R., & Daniel, S. J. (2006). Can virtual reality improve anatomy education? A Randomised controlled study of a computer-generated three-dimensional anatomical ear model. *Medical Education, 40*, 1081–1087.

Nguyen, N., Nelson, A. J., & Wilson, T. D. (2012). Computer visualizations: factors that influence spatial anatomy comprehension. *Anatomical Sciences Education, 5*, 98-108.

Oh, C.-S., Won, H.-S., Kim, K.-J., & Jang, D.-S., (2011). "Digit anatomy": A New technique for learning anatomy using motor memory. *Anatomical Sciences Education, 4*, 132–141. doi:10.1002/ase.223

Pickering, J. D. (2015). Introduction of an anatomy ebook enhances assessment outcomes. *Medical Education, 49*, 522-523

Pickering, J. D., & Bickerdike, S. R. (2016). Medical student use of Facebook to support preparation for anatomy assessments. *Anatomical Sciences Education, 10*, 205-214.

Roach, V. A., Fraser, G. M., Kryklywy, J. H., Mitchell, D. G. V., & Wilson, T. D. (2018). Guiding low spatial ability individuals through visual cueing: The Dual importance of where and when to look. *Anatomical Sciences Education, 12*(1), 32-42. doi:10.1002/ase.1783

Saxena, V., Natarajan, P., O'Sullivan, P. S., & Jain, S. (2008). Effect of the use of instructional anatomy videos on student performance. *Anatomical Sciences Education, 1*, 159– 165.

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

Küçük, S., Kapakin, S., & Göktaş Y. (2016). Learning anatomy via mobile augmented reality: Effects on achievement and cognitive load. *Anatomical Sciences Education, 9*, 411–21.

Kyriazos, T. A. (2018). Applied psychometrics: sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology*, *9*(08), 2207. doi:10.4236/psych.2018.98126

Skinder-Meredith, A. E., (2010). Innovative activities for teaching anatomy of speech production. *Anatomical Sciences Education., 3*, 234–243. doi:0.1002/ase.173

Swinnerton, B. J., Morris, N. P., Hotchkiss, S., & Pickering, J. D. (2017). The Integration of an anatomy massive open online course (MOOC) into a medical anatomy curriculum. *Anatomical Sciences Education, 10*, 53–67.

Tan, S., Hu, A., Wilson, T., Ladak, H., Haase, P., & Fung, K. (2012). Role of a computer- generated three-dimensional laryngeal model in anatomy teaching for advanced learners. *Journal of laryngology and otology*, 126, 395–401.

Trelease, R. B. (2016). From chalkboard, slides, and paper to e-learning: How computing technologies have transformed anatomical sciences education. *Anatomical Sciences Education*, 9, 583–602.

Venkatesh, V., Thong, J. Y. L., Chan, F. K. Y., Hu, P. J.-H., & Brown, S. A. (2011). Extending the two-stage information systems continuance model: Incorporating UTAUT predictors and the role of context. *Information Systems Journal, 21*, 527–555.

Vogt, S., Di Rienzo, F., Collet, C., Collins, A., & Guillot, A. (2013). Multiple roles of motor imagery during action observation. *Frontiers in human neuroscience, 7,* 807. doi:10.3389/fnhum.2013.00807

Weyhe, D., Uslar, V., Weyhe, F., Kaluschke, M., & Zachmann, G., (2018). Immersive anatomy atlas-empirical study investigating the usability of a virtual reality environment as a learning tool for anatomy. *Frontiers in Surgery, 5*, 73. doi:10.3389/fsurg.2018.00073

Wiedenbauer, G., & Jansen-Osmann, P. (2008). Manual training of mental rotation in children. *Learning and instruction, 18*, 30-41.

Yiannoutsou, N., Johnson, R., & Price, S. (2021). Non-visual Virtual Reality: Considerations for the Pedagogical Design of Embodied Mathematical Experiences for Visually Impaired Children. *Educational Technology & Society*, *24* (2), 151–163.

Non-visual Virtual Reality: Considerations for the Pedagogical Design of Embodied Mathematical Experiences for Visually Impaired Children

Nikoleta Yiannoutsou[1], Rose Johnson[2] and Sara Price[2*]
[1]Joint Research Centre, European Commission, Sevilla, Spain // [2]University College London, UK
nyiannoutsou@gmail.com // rosy.johnson@googlemail.com // sara.price@ucl.ac.uk
[*]Corresponding author

ABSTRACT: Digital developments that foreground the sensory body and movement interaction offer new ways of engaging with mathematical ideas. Theories of embodied cognition argue for the important role of sensorimotor interaction in underpinning cognition. For visually impaired children this is particularly promising, since it provides opportunities for grounding mathematical ideas in bodily experience. The use of iVR technologies for visually impaired children is not immediately evident, given the central role of vision in immersive virtual worlds. This paper presents an iterative, design-based case study with visually impaired children to inform the pedagogical design of embodied learning experiences in iVR. Drawing from embodied pedagogy, it explores the process of implementing a classroom-based non-visual VR experience, designed to give visually impaired children an embodied experience of position in terms of Cartesian co-ordinates as they move around a virtual space. Video recordings of interaction combined with feedback from teachers and children contribute to knowledge of iVR learning applications in formal settings by discerning three types of pedagogical practices: creation of a performance space introduction of performative actions and action connected diverse perspectives.

Keywords: Virtual Reality, Visually impaired children, Embodied learning, Cartesian coordinates

1. Introduction

Embodied learning is increasingly important in shaping how to teach mathematics, given evidence suggesting that mathematical knowledge is embodied (Alibali & Nathan, 2012). Embodied learning has particular potential for teaching visually impaired (VI) students, since it engages multiple senses including proprioception, bodily action, touch and hearing, rather than focusing primarily on visual resources. Immersive Virtual Reality (iVR) has the potential to foster embodied forms of teaching and learning mathematics through whole body sensorimotor interaction (e.g., Price et al., 2017), yet the benefits for VI children is not immediately evident, given the central role of vision in immersive virtual worlds. A significant body of work has explored technical application of tactile and multimodal feedback in VR for VI particularly transforming spatial information to auditory cues to support navigation (Cobb & Sharkey, 2007). However, the educational application of iVR with VI children remains under-researched.

Compared to their “seeing” counterparts VI children typically fall behind in science and mathematics by one to three years (Quek et al., 2006). Learning about Cartesian coordinates is important for developing spatial understanding - a challenging competence in everyday life for VI, and considered a fundamental concept for mathematics and science (Knuth, 2000). VR applications lend themselves to spatial concepts like Cartesian coordinates through their potential to support embodied interaction with the learning environment (Johnson-Glenberg, 2018). Yet little work has examined the use of iVR to support VI children’s embodied learning in school settings. Embodied learning has important implications not only for the design of digital environments, but also for the pedagogical design surrounding implementation in the classroom. If the learning of mathematics is shaped by our embodied physical and sensory experience, then the particular physical and sensory experiences we design and how they are pedagogically facilitated are significant in shaping mathematical understanding. iVR has important implications in terms of disrupting existing classroom practices, given different physical space and set up requirements and pedagogical orchestration. The use of iVR with VI children in a school setting, poses additional challenges. While studies referring to iVR in education ground their design and implementation in learning theories like embodied learning (Johnson-Glenberg, 2018) and integrating learning content (Dalgarno & Lee, 2010; Hanson & Shelton, 2008), less attention is given to implementation in classroom contexts, and explicit reference to pedagogy is missing (Southgate, 2020). This is of particular importance given that technology alone does not teach: “it is educational practices that determine how well students learn, and technology is not a process but a tool through which educational practices are mediated” (Rappaport cited in Niederhauser, 2013, p. 249).

Drawing on this perspective our research question asks: What educational practices can support VI children to engage with embodied exploration of mathematical concepts through the use of iVR in the classroom? Thus, our study focuses on informing pedagogical practices that frame the use of iVR to support embodied learning through examining student-facilitator interaction in the iVR environment. We report the iterative development of a pedagogical design to support VI children's learning about Cartesian coordinates with a purposely designed iVR environment (the Cartesian Garden). The Cartesian Garden is made accessible for VI children by focusing on proprioception and bodily movement in space linked to sound, rather than primarily relying on visuals. Through qualitative analysis of interaction and interview data we explore the way VI children used multisensory affordances of the iVR environment for meaning generation to identify pedagogical approaches that enable children to effectively engage with mathematical ideas without visual feedback. We discuss these findings drawing on notions of embodied pedagogy (Nguyen & Larson, 2015) to outline key pedagogical strategies for implementing and supporting a VR embodied learning experience for VI children in the school classroom.

2. Background

2.1. Embodied learning and mathematics

Lakoff and Nunez (2000) assert that mathematical knowledge is understood through the body situated in the physical environment. Evidence for the embodied basis of mathematics has been shown in research with children and teachers (e.g., Alibali & Nathan, 2012; Lakoff & Núñez, 2000), and the critical role of action in underpinning important sensorimotor representations that can be used later in reasoning (e.g., Abrahamson & Sánchez-García, 2016; Nemirovsky et al., 2012). Embodied learning has important implications for the design of technology and the orchestration of the related learning experiences.

Cartesian coordinates have typically been studied with older students who are not visually impaired (e.g., Boyce & Barnes, 2010; Knuth, 2000). The concept of the Cartesian plane is considered critical for spatial perception, spatial relations and navigation, and in understanding mathematical concepts like functions (Knuth, 2000). Cartesian coordinates offer a way of using the environment as reference for navigation and particularly for undifferentiated and large scale environments (e.g., grids on city maps) (Dokic & Pacherie, 2006). For VI children in particular, embodied engagement through patterns of movement have been acknowledged as an instrument for developing spatial perception and integrating spatial information (Jones, 1975). Studies on spatial perception of the blind offer different theories on the role of vision in spatial understanding. While this debate is beyond the scope of this paper, these theories show that the different modalities used for the exploration of space shape the way spatial information is structured. In this paper iVR is designed to support embodied exploration of Cartesian coordinates through sound and sensorimotor interaction with the learning environment.

2.2. Immersive virtual reality and embodied pedagogy

iVR enables users to interact in worlds that simulate aspects or situations of reality that offer new opportunities for education. Most educational VR application studies involve adults at university, college or training contexts (Freina & Ott, 2015). Fewer studies involve middle school students or elementary students, partly due to insufficient knowledge of potential health (e.g., dizziness) and ethical issues involving young children's use of Head Mounted Displays (HMD)), although 5-10 minutes incremental use for young children is reported to be unproblematic (Aubrey et al., 2018).

For the visually impaired, iVR offers suitable navigation training, as it provides a controlled, customizable and safe environment for this purpose. Research in this domain is limited, involving mainly teenagers or adults and focuses on the technical and usability aspects (see Allain et al., 2015). Other research using non-immersive VR with VI children explores the role of sound in mental structure construction and spatial imagery of VI children (Lányi et al., 2006; Sanchez & Lumbreras, 2000). While these studies offer valuable insights into conceptualizations of space by VI children, non-immersive VR studies do not focus on full-body exploration of the environment: an important aspect of spatial thinking. Furthermore, little research has explored the use of iVR in school contexts rather than lab settings with typically developing or VI children. As a result there is lack a focus on pedagogy (Southgate, 2020). While early attempts looked at ways of integrating content (Dalgarno & Lee, 2010; Hanson & Shelton, 2008), more recent studies include introducing iVR technologies to learners and teachers, connection to the curriculum, types of feedback and integration of collaboration problem solving activities (Southgate, 2020; Johnson-Glenberg, 2018). Johnson-Glenberg (2018) proposes guidelines with specific reference to how gestures and body metaphor can be integrated into pedagogical design. However, challenges of implementing iVR in classroom settings and developing appropriate pedagogical strategies has not been explored.

iVR technologies vary in terms of equipment and forms of interaction they enable. Deciding on which type of VR to use has important implications for classroom practice, from cost requirements to physical space, set up, student involvement and interaction, and teaching strategies. HMDs for iVR isolate the user from the physical environment and require space of at least 3X3 meters. The disruptive nature of iVR also poses critical pedagogical implementation challenges in classroom settings, which have not been previously addressed. This paper explores the pedagogical implementation of iVR into the classroom with VI children, specifically attending to the embodied nature of interaction to identify key pedagogical requirements drawing on notions of embodied pedagogy.

Embodied pedagogy focuses on how to engage the body and space in learning (Nguyen & Larson, 2015). It is defined "as learning that joins body and mind in a physical and mental act of knowledge construction. This union entails thoughtful awareness of body, space, and social context" (Nguyen & Larson, 2015, p. 332). In embodied pedagogy, knowledge construction takes place in a performance space which prescribes spatial arrangements not only of objects but also of body position, orientation, movement and action. The relationship between body and mind is expanded to include the material situation of the performance space and the human (i.e., teacher- student, student -student) interdependencies (Southgate & Smith, 2016). Meaning construction in embodied pedagogy involves a cyclical process of mindful action and reflection which shapes the learner and the space and spatial arrangements where action takes place. Action is not simply situated in the environment. Instead people and the environment, transform and shape each other through ongoing transactions (Sund et al., 2019). While these transactions are instrumental for progress, introducing diverse perspectives in the action–reflection cycle can instill the physical in the reflective and the reflective in the physical in unexpected ways, resulting in new modes of learning (Southgate & Smith, 2016). iVR technology foregrounds space and physical action in the teaching and learning process, thus demanding engagement with embodied pedagogy. Our study draws on principles of embodied pedagogy to identify particular considerations that can inform the implementation of iVR in the VI classroom to effectively structure an embodied learning experience.

3. The Cartesian garden environment

The environment uses HTC Vive to create a VR experience space measuring 5mx5m, consisting of a floor-based virtual grid in the form of a virtual garden, with flowers at each of the coordinates. X and y-axes run from 0-5 in the standard layout for Cartesian coordinates for children. Children walk along the axes to find and collect target flowers at specified coordinates. For VI children sounds and vibration on the controller provide feedback that maps movement to coordinates. As the child moves around the grid, they hear sounds from a set of speakers that indicate their position in relation to the grid lines. A controller attached to the child's waist using a belt vibrates when the child encounters a cross point. To find out their exact position (in relation to the coordinates) the child presses the controller trigger and the position is defined using sound.

Sounds: Sounds convey directional and mathematical properties: movement along the x-axis is indicated by flute notes and movement along the y-axis is indicated by violin notes. As the child walks along the x-axis they hear only flute notes, when they walk along the y-axis they hear only violin notes. The directional properties of the sound involve embodied action: the further the child moves from the origin on one axis the more notes they hear. In order to hear a combination of notes they need to move within the middle of the garden (within the quadrant). Thus, position (3, 2) is expressed by three flute notes, followed by two violin notes. Two different instruments played sequentially and mapped to the two axes aims to demonstrate that coordinates are pairs of numbers describing distance in orthogonal directions. An ambient sound indicates that the child is on a grid line, but stops when the child is off the line, a "ding" sounds when they reach a cross point, and white noise is heard when they are on point 0 (1,0 would be one flute note followed by a white noise).

Screen: A 2D computer screen depicts the garden as a grid of flowers with the child's position and orientation indicated by a red triangle (Figure 1). In this study the screen was used by researchers in order to identify technical problems, like lack of feedback.

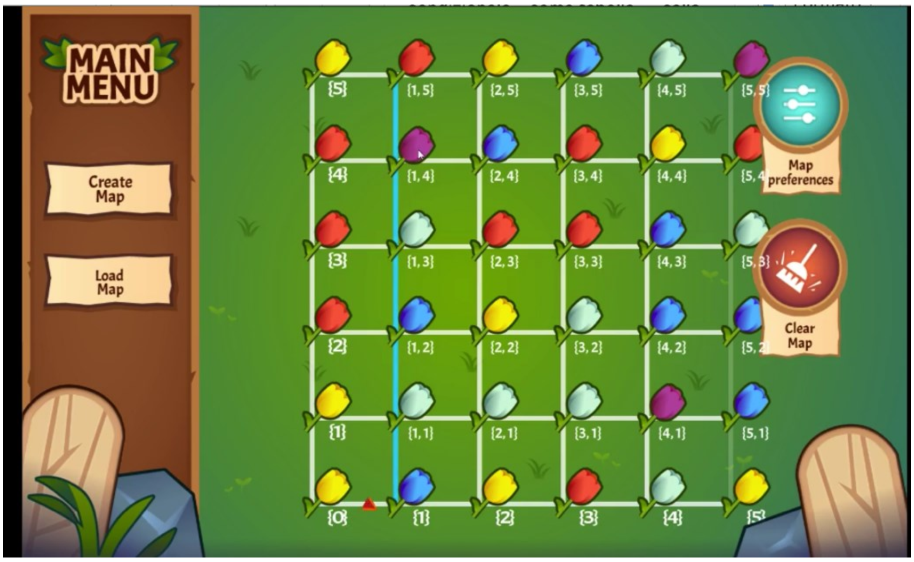


Figure 1. The computer screen shows the garden. The red triangle near (1,0) is the position of the player

VR Headset: When wearing the VR headset children with usable vision are immersed in a garden of flowers. However, for VI children relying on other modalities the VR headset was not used, position being conveyed by sound and vibration as described above.

4. Methodological approach

An iterative design-based research approach was taken to develop an effective pedagogical design around the VR experience with VI children, drawing on researcher and teacher facilitation of child interaction. Design based research aims to improve educational practice through iterative design, development and deployment of educational activities in real world settings, developed through researcher-practitioner collaboration, with a view to generating design guidelines (Wang & Hannafin, 2005). In this paper we report on a two-phase study focusing on the iterative design of the pedagogical strategies needed to effectively implement a iVR learning experience for VI children into the school classroom. The study drew on children's interaction, researcher-teacher facilitation, teacher and child interviews. Informed consent to participate was obtained from children and their parents. Methods for each phase are detailed under each study.

5. Phase 1

Phase 1 aimed to identify a set of pedagogical designs to be implemented and evaluated in Phase 2. Phase 1 was conducted with one blind child aged 8 [C2]. The child interacted with the iVR experience with the help of a researcher and his teacher for one hour. This was followed by an informal interview with the child about his experience, and with the teacher about pedagogical feedback. A second researcher video recorded the session and took notes while a third operated the technology. The child was encouraged to comment on his experience and was told that this would help us identify if something was not clear or difficult to understand, to then improve the experience. Data comprised video recording of interaction and observation notes, including the discussion with teacher and child. Researcher reflection on these data led to identifying key pedagogical design issues around the VR experience. Three key aspects emerged that informed development of the pedagogical design for Phase 2: space, body movement and concept/activity framing.

5.1. Space: Physical and virtual

Physical Space: played a critical role as the way children moved in the space was essential for fostering meaningful embodied engagement with Cartesian coordinates. The research sessions took place in a dedicated classroom with free space of 5X5 for the VR overlay, within which children could freely move. This room was unfamiliar to the children: it was the science classroom which children had visited only twice, and the furniture had to be rearranged to meet requirements for the VR set up. Thus, we needed to orientate them, explain that they would only be moving in the free space between the door and the window, and assure them that the researcher would be next to them to guide and make sure they did not bump into anything during the interaction.

VR Interaction Space: Aligning the VR overlay in the physical space and 2D representation was found to be critical since the researcher and teacher needed to easily identify where the child was in the virtual garden, to guide them appropriately and ensure they received feedback on their position. Sticky tape was used to mark key elements of the Cartesian plane (i.e., axes, origin, and distance between flowers on the x and y axes) on the physical floor space. Researchers used floor markings to guide the child to refine his movement appropriately. However, C2 required more support within the central parts of the quadrant to effectively receive sound and haptic feedback. The floor marking was thus extended to include the cross point positions with an emphasis on the position of the child's feet from the VR feed (green cross) and the position of the controller (blue line), which were not directly aligned because the controller sits fractionally in front of the feet (Figure 2). This alignment was critical in facilitating the transactional relation between children (moving in space), the researcher-teacher (supporting children to refine their movement) and the iVR environment (offering feedback).

Figure 2. The physical space marked out to facilitate VI children's movement in the VR space, aligned to the grid

5.2. Movement

Phase 1 revealed key difficulties for C2 in mapping their bodily movement to the grid. Moving through the Cartesian Garden was not only dependent on the child's understanding of their current position and the position they were aiming for, but also their bodily orientation with respect to the axes. For VI children there are few external cues to inform them in which direction they are facing. The iVR experience gives feedback on position but not orientation. While C2 made a quarter turn with his body every time he changed from moving in the x direction to moving in the y direction, it was not always clear which direction to turn. In addition, it was not always easy for him to 'sense' when he had achieved a precise 90-degree turn. To avoid disorientation the researcher had to specifically state the direction the child was facing. As the session progressed C2 adopted a fixed orientation with his body always facing in the y direction (as if looking at a graph on a piece of paper) and stepping sideways along the x-axis. In this way, the child used their body as an anchor for orientation. Following teacher advice, the VI children were led by the hand from the front, guiding them so that the researcher steps into the unknown first, to help children feel safe, yet allow them to follow at their own pace.

5.3. Concept/activity framing

Phase 1 revealed the need to familiarise VI children with the notion of grids. C2 was not easily able to make connections between his whole-body experience and the external layout of the garden. During interaction it became apparent that he was not familiar with the concept of a grid. The teacher intervened and introduced a tactile grid to help explain the idea to the child:

Now you find those squares, if you follow that line all the way up (she guides his hands) and there's another line that goes across (she guides his hands) So when we are talking about the garden, the garden kind of looks like this. So, this garden that you are walking around on.

Non-VI children using a headset could see the grid and explore ways of moving related to sound and visuals. For VI children reduced access to the visual representation, demands alternative ways of introducing the key grid structure and related movement. By combining the tactile grid and the fixed body orientation strategy C2 was better able to negotiate the space and pay attention to how the sounds related to his location and could guide his next steps.

Phase 1 also highlighted challenges in explaining the garden and linking mathematical terminology. Although we introduced axes as lines following a particular direction and encouraged C2 to explore how the sound changed as he moved along the axis, he had difficulty identifying the two different sets of sounds mapped. This highlighted the need to develop a strategy that clearly explained the mapping of sounds to two different paths.

5.4. Adaptations for Phase 2

Based on Phase 1 the following changes were made to the experience structure and pedagogical process:
Awareness of Physical Space: Due to children's unfamiliarity with the physical space we identified the room they were in and described the layout to provide the necessary information for children to locate themselves in space.

VR Interaction Space: The physical space was marked with paper tape to enable researchers to facilitate interaction with the VR experience reducing the difficulties in finding the exact location where the system provided positional feedback to the children.

Tactile grid: An improved tactile grid (Figure 3) was included to enable children to explore the layout of the grid and spatial relations between different positions before they started moving in the garden space.

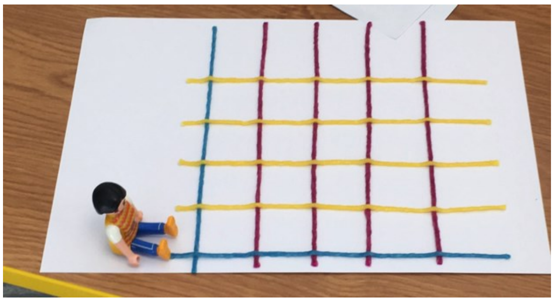


Figure 3. The tactile grid made from wax string pressed on to an A4 sheet of card, and a playmobile figure

After their interaction in the iVR space, children were presented with a playmobile figure and asked to move it on the tactile grid to specific coordinates (Figure 3), in order to explore any learning transfer from the physical VR space to a more abstract 2D space.

Guiding: The researcher guided children by the hand from the front to ensure that they remained within the iVR experience space. However, to avoid implicitly leading the child to the correct answer the researcher kept some distance between them, allowing the child to move towards the direction indicated by the child.

Fixed Body Orientation: The children were advised to stand facing towards the y-axis and take a combination of side-steps and forward or backward steps to navigate the grid, as follows: "In this garden you have to walk in a specific way: first you have to move sideways as many squares as the flute player tells you, and then you have to move forward as many squares as the violin player says."

Violin and Flute Player: To help children link the different sounds to location along the axes we explained that a flute player was sitting at the end of one path (x-axis) and a violin player at the end of the other (y-axis), and as they walk towards the musician, they will hear more notes. We renamed the origin to "garden gate" and spoke about position in terms of number of flute and violin notes, rather than coordinates at this stage.

Magic Flowers: To better integrate the flowers and give a more play-like quality to the experience, we introduced the idea of a magic flower: two flutes, three violins, which the children had to find. If the children found the magic- flower they could choose the magic powers it would give them.

6. Phase 2

Phase 2 implemented the changes in the pedagogical design outlined above and explored how these appropriations fostered embodied engagement with Cartesian coordinates mediated by the iVR experience.

6.1. Participants

Seven visually impaired children (2 girls and 5 boys) aged between 8 and 11 years took part, including the child from Phase 1. The teacher classified the children according to their visual ability into two main groups: (a) children who could see shapes and/or enlarged prints if brought close to their face and could read extremely large lettering. These children moved around without the need of a cane (1 girl and 4 boys); (b) children who were blind and needed a cane to move (one boy and one girl). Within these two broad categories there were differentiations including children who had lost their sight after birth and children who were born blind or with visual impairment. This differentiation is important since children who are born visually impaired or blind have very poor or no visual representations, impacting how they understand space and instructions. Given the range of visual impairments, ages and educational needs, each child took part in the experience individually.

6.2. iVR Interaction sessions

The researchers visited the classroom prior to the research sessions, introduced themselves, explained the activities and the technology. Children had the opportunity to touch and explore the controllers and belt in order to familiarise themselves with the equipment, enabling children to demystify the technologies and reduce potential anxiety with unknown things (people, tasks, technologies). Each session lasted approximately 45-50 minutes. Each child was accompanied by the teacher, who advocated on behalf of the child in case they became tired or needed to stop. The sessions were structured in three parts: a short informal interview to identify children's experience with computer games, and knowledge of grids and coordinates; introduction to the Cartesian plane through exploration of a tactile grid (Figure 3); a short discussion about their experience and completion of tactile grid tasks. One researcher took the role of facilitator, a second took notes and a third operated the technology and video camera. Most participants had not learned about grids before.

6.3. Data collected and analytical approach

Data included video recordings of children's interaction and pre and post interaction discussions, screen capture of the 2D representation of children's movement, and researcher reflective notes of the sessions, including interactions with the teacher. Video data transcription attended to movement in relation to the physical space, the virtual grid and technology feedback, and verbal interactions. The analysis follows an inductive (looking for emerging themes) and deductive (informed by concepts of embodied pedagogy) process to derive key implications for the development of the pedagogical design towards embodied exploration of the Cartesian Plane. The following section presents the key findings used as a basis to evaluate our pedagogical design. While each section offers insights from all participants, key illustrative examples are drawn from four cases.

7. Findings and discussion

Phase 2 enabled explored the role of the implemented pedagogical strategies in supporting VI students' embodied exploration of mathematics. The findings show how bodily movement and positioning can effectively foster VI children's engagement with the Cartesian plane, and the role that sound and researcher facilitation played in this process. These findings are discussed within the frame of embodied pedagogy (Nguyen & Larson, 2015). The iVR design itself creates a performative space. Our findings illustrate how populating this space with performative actions and introducing diverse perspectives (Nguyen & Larson, 2015) facilitated mindful action i.e., embodied action and reflection.

7.1. Performance space

The Cartesian Garden creates a "performance space" through: (a) the physical space allowing for whole body movement; (b) the virtual overlay that provides a narrative for interaction; and (c) the relationship between body and physical-virtual space. Phase 1 analysis showed that spatial arrangements enabling children to know the relative positioning and orientation of their bodies was critical for interaction in the iVR environment. These spatial arrangements formed the basis of performative actions that supported embodied learning interactions.

The narrative of the magic garden provided meaning to the task and raised students' interest through the introduction of make-believe elements (garden of flowers) and set the scene for identifying objects, bodies, relations and actions in the specific space (i.e., collecting flowers from a regulated garden layout). The pedagogical appropriations included the introduction of a set of performative actions which embodied key aspects of the Cartesian Plane (axes, orientation, order of coordinates), aiming to facilitate embodied interaction with the mathematical concepts.

7.2. Performative actions

Embodied pedagogy acknowledges performative thinking as an embodied process linked mainly to role play (Nguyen & Larson, 2015). In our study, performative thinking is connected to specific appropriations of the learning environment, that allow performative actions or ways of moving in space which embody specific mathematical ideas. These performative actions took the following forms:

7.2.1. Musical paths: Embodying properties of the Cartesian plane

In the iVR environment, walking along the axes was linked to the sound of different instruments. The task descriptions introduced a narrative involving two musicians sitting at the end of each path (axis). This added directional and distance properties to the musical metaphor: the closer the child moved towards the musician the more notes they could hear. The sounds thus shaped a set of performative actions, which in this case involved ways of moving in space. It is worth noting that these performative actions, which combine movement with sound constitute a common practice in the sociocultural context of VI children.

When children were introduced to the garden they began at the origin (0,0 or the garden gate) with their body oriented to face the y-axis direction. In Phase 2 children were shown the path to the flute player (x-axis) first, and were told to take side steps along the axis. As they stepped along the x-axis they experienced the flute sounds at each cross point, the number of flute notes increasing as they went further along the path. They were then taken back to the "garden gate" and stepped forward along the y axis, experiencing the path to the violinist, again hearing an increased number of violin notes at each point. Once they reached 4 violin notes (0,4) they were asked "if I wanted to hear flute notes as well as violin notes which way should I step?" This was a transition point where they had to go from perceiving flute and violin (x and y coordinates) as describing the paths of each music player (axes) to perceiving them as locations requiring movement in a direction parallel to one axis. Few children were able to immediately work out that they needed to step sideways into the middle of the garden to hear the flute as well as the violin. One child (C3), who was partially sighted, described how she did map this use of sounds:

"*Each path has its own instrument, for example that's the violin and that's the flute. So if I'm at the top of the violin path I just have to go here [she indicates walking parallel to the x axis at y = 4] I don't have to go back here to the actual path because the path is invisible on the floor but it's still the same way as the violin line goes – the same way as the flute line [correcting herself] - So I just go the same way as the flute line rather than going back down, otherwise I won't get the flower.*"

Children who struggled with this initially only walked along either of the paths they had been shown (x and y axis), requiring encouragement to explore the middle space. Even after being shown that there were positions in the middle of the garden, with both flute and violin notes, some did not immediately move into this space. For example, although C5 had already explored the middle of the garden, he still needed prompting to step off the axes to find coordinates in the middle of the quadrant.

[C5 is looking for two flutes and four violins. C5 has identified that he needs to move right to get the flutes.]
...
R1: So we have two flutes, now we need four violins which way do we go?
[C5 moves back towards the origin to start from there up the y axis. C5 gradually moves up the y-axis to (0,4).]
R1: How many violins do we have?
C5: Four
R1: But how many flutes do we need?
C5: Two
R1: Where are we going to get them?
[C5 gestures back to the origin]
R1: If we move along the first path, we will only have the flutes and no violins. But we need 2 flutes and 4 violins so...

C5: This way [gestures towards the middle of the garden]

This movement pattern enabled children to listen to the sounds sequentially: first they heard one sound, moved back to the origin and then in the other direction to hear the second sound. This pattern of moving towards one music player, then return to the origin and walk on the other axis towards the second instrument player, served as an embodied manifestation of children's thinking, allowing for reflection and teacher intervention. Later when C5 was asked to take the playmobile figure to position (3,2) on the grid, he also moved the figure to (3,0) first, back to the origin and then to (0,2). While children were prompted to explore the middle of the garden to hear both sounds, four were able to do this effectively. This suggests the need for additional strategies to help them conceptualise the paths that run parallel to the axes.

7.2.2. Orientation and movement: Embodying properties of Cartesian coordinates

Children were guided to adopt a fixed body orientation facing the y-axis, to step sideways to move along the x-axis and forward to move along the y-axis. This unorthodox way of "walking in space" employs the idea of congruence between embodied action and concept (Johnson-Glenberg, 2018) to integrate key properties of Cartesian coordinates involving specific rules for interaction (i.e., x then y), and orientational alignment with other representations e.g., tactile grid. Children were asked to find a magic flower positionally described as 3 flute 4 violin (coordinates (3,4)). Some children started from the origin (0,0) walking sideways along the x-axis then moving forward along the y-axis (Figure 4), others adopted the pattern: x, origin, then y, outlined above (7.2.1). Children used the controllers to get sound feedback about their position. The extract below demonstrates how C4 applied the sideways first then forward pattern of movement in combination with sound feedback in order to successfully find the magic flower using coordinates.

R3: Magic flower at three flute, three violin.
…
[C4 takes small step forwards, hears ding, presses trigger, two flutes.]
R1: How many were there before?
C4: Two
…
[C4 takes big step to right then little step forward, hears ding, presses trigger, three flutes]
R1: Great, so we've got three flute and [want] three violins.
R2: So, where do you go from there to get three violins?
…
[C4 takes a big step forwards, ding, presses trigger, three flute, one violin. Takes another big step forwards, ding, presses trigger, three flute two violin. Takes another big step forward, ding, presses trigger, 3 flutes, 3 violins]

In this process C4 used the controller to get sound-based positional feedback and adjusted his movement accordingly (i.e., calculated the direction to move and how many steps to take). All children used the feedback offered by the controllers but two out of the seven children waited for researcher feedback rather than interpreting feedback from the controllers. The researchers prompted them to listen to the feedback carefully, verbalise what they were hearing, and describe how they should move. In the iVR environment, feedback facilitates the mindful action–reflection cycle of embodied pedagogy: the sound feedback helps the child to become aware of their position in the performance space, to reflect on this position and decide their next action. The combination of feedback with introduction of the specific pattern of movement proved instrumental in supporting embodied learning for VI in the iVR experience.

7.2.3. Guidance: Embodied instruction

The interaction between the teacher/researcher and student generated a set of performative actions specific to the iVR space. All 7 children needed guidance to remain within the performative space given the small unbounded space aligned to VR. The teachers-researchers adopted specific ways of guiding the children, informed by the sociocultural context of the students and school. This involved leading them by the hand from the front, but letting the child indicate their chosen direction of movement, or giving directional hints through leading when necessary. In this way guidance took the form of embodied instruction and intervention.

7.3. Diverse perspectives: Multisensory experiences

Diverse perspectives were introduced by supporting different and interconnected sensory modalities. Sound feedback was connected and responsive to movement designed to enable reflection informing mindful action. Further, different activities on the Cartesian Plane were introduced, placing the body in different points of view of the same concept, and creating embodied links across representations (Nguyen & Larson, 2015). Specifically, the iVR environment offered a space where the child's body became part of the grid, facilitating ego-centric exploration; the tactile grid used before the activity enabled children to haptically gain an introductory overview (similar to visually perceiving a paper-based grid) of the Cartesian plane as a space consisting of squares with intersections as key points for describing position; in the post interaction discussion, the grid offered an allocentric perspective when children were asked to move a playmobile figure into a position described by coordinates.

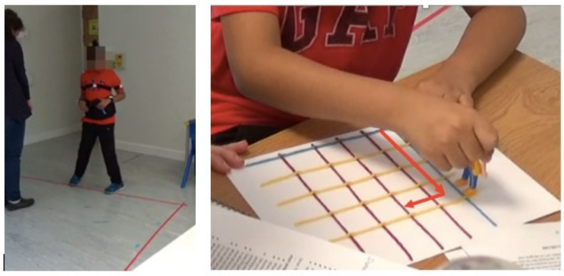


Figure 4. Patterns of movement transferred from the 3D space to the 2D representation of the Cartesian plane

In this study performative actions played a role in transcending different perspectives and supporting transfer of knowledge from one representation to the other. In the post interaction discussion all children were able to transfer their knowledge from the physical space to the 2D haptic grid representation, using the same moving pattern as a common referent between the two spaces. Four children were able to find the correct position in the Cartesian plane. Three children using on the haptic grid their patterns of movement in space, focused only on the two axes (x,y) without considering coordinates in the middle of the grid (see section 7.2.1). The powerful role of patterns of movement is illustrated in the post interaction discussion with C4. The marking of the trajectory on Figure 4 shows that C4 moved first sideways along the y-axis (rather than sideways along the x-axis), allocating the first coordinate to the y instead of x-axis. The original orientation of the playmobile figure was changed by the child - during random play - to face towards the x-axis, thus influencing the direction of movement. This finding is interesting, if we consider previous research with typically developing children (8-9 years old) which demonstrated that this transition from enactment in the physical space to using the Cartesian plane on paper was challenging (Price et al., 2020).

7.4. Child reflections

The post interaction discussion involved the evaluation of learning through the use of the tactile grid, which was mentioned in the sections above, but also focused on exploring children's perception of the experience, and their evaluation of key functionalities (like sound feedback). While all children found the experience interesting and engaging, this can be attributed to factors like the novelty of technology, deviation from school routine etc. Further, children reflected on key aspects of the experience like the sound feedback. The majority of the children mentioned some difficulty in distinguishing the two sounds related to the x and y axes. This technical aspect having also conceptual implications as shown in the sections above, demonstrated the need for refining the sound feedback.

8. Conclusion: Pedagogical appropriation of iVR for VI

This paper shows the importance of developing appropriate embodied pedagogical designs to support effective integration of iVR in school settings for VI children. Through an iterative design process, our study offers insights into the prerequisites for implementation of iVR in the classroom and embodied pedagogical appropriations to support VI children to explore the concept of cartesian coordinates.

8.1. Prerequisites for classroom implementation

iVR disrupts existing school practices, calling for requirement specifications to implement the Cartesian Garden experience. Firstly, our study revealed the need for specific spatial arrangements and for appropriate physical space devoted to the iVR activity. Identifying specific space to deploy full tracking iVR systems highlights the spatial constraints of schools built for an industrial era (Southgate et al., 2019). Secondly, orchestration of the experience allows one child per session due to space requirements and VR features which foster individual interaction. In addition, at least one teacher per session is needed to facilitate the activity and provide appropriate support for the student, and potentially two adults with one to monitor/manage the technology, since iVR requires constant surveillance (Southgate et al., 2019). While individual sessions with children in VI schools are not unusual, iVR currently has a considerable financial cost and requires many human resources. These observations are relevant not only to educational practitioners but also technology designers.

8.2. Pedagogical appropriations of iVR to support embodied learning

Our study, framed by embodied pedagogy, enriches existing recommendations specifically addressing the use of iVR in classroom settings (Johnson-Glenberg, 2018; Southgate, 2020) by discerning the role of performance space, diverse perspectives and performative actions as key elements in supporting embodied learning. Table 1 presents a summary of pedagogical appropriations their connection to concepts of embodied pedagogy and their pedagogical aim.

Table 1. Appropriations of iVR to support embodied pedagogy of mathematics learning

Embodied pedagogy	Appropriations	Function
Creating a performance space	Marking the floor to map the physical to the virtual to support adult facilitation	Facilitate the transactional relation between teacher-researcher, student, space
	Orientation alignment between computer screen and physical space.	Facilitate the transactional relation between teacher-researcher, student, space
	Introducing narrative or make-believe elements (magic flower)	Make the task meaningful and interesting for the children, introduce performative thinking
Performative actions	Familiarisation with performative space (verbal descriptions of space); Guidance strategies (from the front); Interaction strategies (moving/ using the controller to get feedback)	Facilitate the transactional relation between teacher-researcher, student and the environment
	Fixed body orientation mapped to axes travelling in orthogonal directions. Specific movement patterns: step sideways (along x axis) followed by forwards (along y axis) Musical paths metaphor to map to axes, providing information relating to direction, distance and position	Integration of key mathematical concepts in embodied actions
Diverse perspectives	Connect different perspectives: allocentric (haptic grid with playmobil figure) & body-centric perspectives (iVR through enacting position) connected with the same performative actions (i.e., fixed orientation and movement pattern) applied in both.	Support students to use performative actions as an embodied means to establish connections between different representations.

The appropriations above involve iVR designed for open embodied interactive exploration of mathematics. Drawing from embodied pedagogy our interventions fall into three key categories: performative space, performative actions and action connected diverse perspectives. Performance space includes the objects, persons, relations and their roles. In our intervention we populated the performance space with a narrative giving meaning to objects, persons and space, and with anchors (marks on the floor or orientation of the screen) that facilitated interactions between children, teacher-researcher and space. The anchors were implemented to complement the feedback provided by the iVR experience, with key additional information. Our study brings the concept of performative actions into educational practices. These are prescribed actions in the environment (objects, relations, people, space), which when enacted embody elements of the concepts under exploration (i.e., action congruence, Johnson-Glenberg, 2018). Student modifications of some of these movements demonstrated how they embodied the concepts of Cartesian coordinates, sometimes manifesting misunderstandings. Diverse perspectives implemented through role-play or through the use of different representations is an important aspect of pedagogy. Our intervention demonstrated how the use of resources, which allow children to implement the same performative actions across different representations, has the potential to support transfer of knowledge. While some of these appropriations could be integrated into the iVR design (e.g., movement patterns), a focus on performative actions, performance space and action connected to diverse perspectives can also inform teacher interventions and pedagogical design around the use of iVR.

Acknowledgements

This research was part of the weDRAW project, which received funding from the European Union's Horizon 2020 Research and Innovation Programme, Grant Agreement No. 732391. The adapted iVR environment for VI children was developed by Ignition Factory in collaboration with Casa Paganini, IIT, Instituto Chiossone in Italy. We would like to thank the wider project team, the school, children and teachers involved in this study.

References

Abrahamson, D., & Sánchez-García, R. (2016). Learning is moving in new ways: The Ecological dynamics of mathematics education. *Journal of the Learning Sciences, 25*(2), 203–239. doi:10.1080/10508406.2016.1143370

Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *Journal of the Learning Sciences, 21*(2), 247–286. doi:10.1080/10508406.2011.611446

Allain, K., Dado, B., Van Gelderen, M., Hokke, O., Oliveira, M., Bidarra, R., Gaubitch, N. D., Hendriks, R. C., & Kybartas, B. (2015). An Audio game for training navigation skills of blind children. In *2015 IEEE 2nd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)* (pp. 1–4). doi:10.1109/SIVE.2015.7361292

Aubrey, J. S., Robb, M. B., Bailey, J., & Bailenson, J. (2018). *Virtual reality 101: What you need to know about kids and VR*. Common Sense.

Boyce, A., & Barnes, T. (2010). BeadLoom Game: Using game elements to increase motivation and learning. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games - FDG '10* (pp. 25–31). doi:10.1145/1822348.1822352

Cobb, S. V. G., & Sharkey, P. M. (2007). A Decade of research and development in disability, virtual reality and associated technologies: Review of ICDVRAT 1996-2006. *The International Journal of Virtual Reality* 6(2) 61-68

Dalgarno, B., & Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology, 41*(1), 10–32. doi:10.1111/j.1467-8535.2009.01038.x

Dokic, J., & Pacherie, E. (2006). On the very idea of a frame of reference. In M. Hickmann & S. Robert (Eds.), *Typological Studies in Language* (Vol. 66, pp. 259–280). doi:10.1075/tsl.66.16dok

Freina, L., & Ott, M. (2015). A Literature review on immersive virtual reality in education. *State Of The Art and Perspectives, 1*, 133–141. doi:10.1007/BF00398472

Hanson, K., & Shelton, B. E. (2008). Design and development of virtual reality: Analysis of challenges faced by educators. *Educational Technology & Society, 11*(1), 118-131.

Johnson-Glenberg, M. C. (2018). Immersive VR and education: Embodied design principles that include gesture and hand controls. *Frontiers in Robotics and AI, 5*. doi:10.3389/frobt.2018.00081

Jones, B. (1975). Spatial perception in the blind. *British Journal of Psychology, 66*(4), 461–472. doi:10.1111/j.2044-8295.1975.tb01481.x

Knuth, J. E. (2000). Student understanding of the cartesian connection: An Exploratory study. *Journal of Research in Mathematics Education, 31*(4), 500–507. doi:10.2307/749655

Lakoff, G., & Núñez, R. E. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York, NY: Basic Books, Inc.

Lányi, C. S., Geiszt, Z., Károlyi, P., Tilinger, Á., & Magyar, V. (2006). Virtual reality in special needs early education. *The International Journal of Virtual Reality,* 5(4) 55-68.

Nemirovsky, R., Rasmussen, C., Sweeney, G., & Wawro, M. (2012). When the classroom floor becomes the complex plane: Addition and multiplication as ways of bodily navigation. *Journal of the Learning Sciences, 21*(2), 287–323. doi:10.1080/10508406.2011.611445

Nguyen, D. J., & Larson, J. B. (2015). Don't forget about the body: Exploring the curricular possibilities of embodied pedagogy. *Innovative Higher Education, 40*(4), 331–344. doi:10.1007/s10755-015-9319-6

Niederhauser, D. S. (2013). Learning from technology or learning with technology. In M. P. Clough, J. K. Olson, & D. S. Niederhauser (Eds.), *The Nature of Technology* (pp. 249–267). doi:10.1007/978-94-6209-269-3_14

Price, S., Duffy, S., & Gori, M. (2017). Developing a pedagogical framework for designing a multisensory serious gaming environment. In *Proceedings of the 1st ACM SIGCHI International Workshop on Multimodal Interaction for Education - MIE 2017* (pp. 1–9). doi:10.1145/3139513.3139517

Price, S., Yiannoutsou, N., & Vezzoli, Y. (2020). Making the Body Tangible: Elementary Geometry Learning through VR. *Digital Experiences in Mathematics Education, 6*, 213-232. doi:10.1007/s40751-020-00071-7

Quek, F., McNeill, D., & Oliveira, F. (2006). Enabling multimodal communications for enhancing the ability of learning for the visually impaired. In *Proceedings of the 8th International Conference on Multimodal Interfaces - ICMI '06* (pp. 326). doi:10.1145/1180995.1181056

Sanchez, J., & Lumbreras, M. (2000). Usability and cognitive impact of the interaction with 3D virtual interactive acoustic environments by blind children: (705372011-033) [Data set]. *American Psychological Association*. doi:10.1037/e705372011-033

Southgate, E. (2020). *Virtual reality in curriculum and pedagogy evidence from secondary classrooms* (1st ed.). New York, NY: Routledge, Taylor & Francis Group.

Southgate, E., & Smith, S. P. (2016). Pedagogical theory and embodiment: Some provocations for virtual and augmented reality in education. In *IEEE Virtual Reality 2016*, Greenville, South Carolina.

Southgate, E., Smith, S. P., Cividino, C., Saxby, S., Kilham, J., Eather, G., Scevak, J., Summerville, D., Buchanan, R., & Bergin, C. (2019). Embedding immersive virtual reality in classrooms: Ethical, organisational and educational lessons in bridging research and practice. *International Journal of Child-Computer Interaction, 19*, 19–29. doi:10.1016/j.ijcci.2018.10.002

Sund, L., Quennerstedt, M., Öhman, M., & Hoadley, U. (2019). The Embodied social studies classroom – Repositioning the body in the social sciences in school. *Cogent Education, 6*(1). doi:10.1080/2331186X.2019.1569350

Wang, F. & Hannafin, M.J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, 53, 5-23.