The Journal of Educational Technology & Society has Impact Factor 2.133 and 5-Year impact factor 2.682 according to Thomson Scientific 2018 Journal Citations Report.

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

Oct. 2020 Journal of Educational Technology & Society

vol.23 no.4

Journal of

Educational Technology & Society

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

Volume 23 Issue 4

October 2020

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

Educational Technology & Society
An International Journal

Aims and Scope
The journal of *Educational Technology & Society* (ET&S) is an open-access academic journal published quarterly (January, April, July, and October) since October 1998. 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.

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

Publisher
International Forum of Educational Technology & Society

Host
National Yunlin University of Science and Technology, Taiwan

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

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

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

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

Guidelines for authors
Submissions are invited in the following categories:

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

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

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

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

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

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

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

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

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

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

Educational Technology & Society
Volume 23 Number 4 2020

Table of contents

Full Length Articles

Huang, H.-W., King, J.-T., & Lee, C.-L. (2020). The New Science of Learning: Using the Power and Potential of the Brain to Inform Digital Learning. *Educational Technology & Society*, *23* (4), 1–13.

The New Science of Learning: Using the Power and Potential of the Brain to Inform Digital Learning

Hsu-Wen Huang[1*], Jung-Tai King[2] and Chia-Lin Lee[3,4,5,6]
[1]Department of Linguistics and Translation, City University of Hong Kong, Hong Kong // [2]Institute of Neuroscience, National Yang-Ming University, Taiwan // [3]Graduate Institute of Linguistics, National Taiwan University, Taiwan // [4] Department of Psychology, National Taiwan University, Taiwan // [5]Graduate Institute of Brain and Mind Sciences, National Taiwan University, Taiwan // [6]Neurobiology and Cognitive Neuroscience Center, National Taiwan University, Taiwan // hwhuang@cityu.edu.hk // pax2@nctu.edu.tw // chialinlee@ntu.edu.tw
[*]Corresponding author

(Submitted June 2, 2020; Revised July 1, 2020; Accepted August 12, 2020)

ABSTRACT: Integrating education practices and measurements of brain activity has the potential to make learning more engaging and productive. Direct recordings of electrical activity in the brain provide important information about the complex dynamics of the cognitive processes and mental states that occur during learning, which can ultimately empower learners. In this article, electroencephalographic (EEG) methodologies, including the time-frequency and event-related potential techniques, are introduced, and the application of these techniques to studies of digital learning studies is discussed. Considerations of how to collect high quality data in both laboratory and real world settings are also presented, along with potential research directions. Finally, a general guideline for publishing results is offered. These issues are critical for producing useful applications of EEG studies to the digital learning research community.

Keywords: Digital learning, Electroencephalograph (EEG), Event-related potentials, Dry-wireless EEG

1. Digital learning with a cognitive neuroscience approach

The emerging interdisciplinary field of learning science applies empirical research to educational designs with the aim of improving learning processes and outcomes (Fischer, Hmelo-Silver, Goldman, & Reimann, 2018). There is a long history of the shaping of educational practices based on knowledge of how the human mind encapsulates information and interacts with learning contexts (Vygotsky, 1978). However, because the flow of information is mostly unidirectional when using traditional styles of learning with textbooks, understanding of the effects of individual differences among learners and of the interactions between learners and learning contexts during the knowledge acquisition process is limited. The recent emergence of evidence-based digital learning provides opportunities to improve the effectiveness of learning for a wide range of learners and promote positive interactions between learners and learning contexts (Lan, 2020; Wu, Lan, Huang, & Lin, 2019). However, most studies have been based on behavioral findings, and evidence from brain activity data is relatively limited. To produce high quality studies and useful applications with brain activity data, the theoretical underpinnings of cognitive neuroscience and educational psychology must be considered.

Among cognitive neuroscience methodologies, electroencephalography (EEG) has been used to assess learners' spontaneous brain electrical responses. In what follows, we briefly describe the methodology, introduce the time-frequency and event-related potential (ERP) techniques and review some examples of how they are applied in digital learning studies. We then offer considerations about how to collect high quality data, and we provide suggestions for potential research directions. Finally, a brief guideline for publishing results is presented.

1.1. Why is EEG useful for learning science?

The 100-year history of EEG affords a rich and diverse spectrum of applications and provided solid foundations for research in a wide variety of fields. However, EEG offers a particularly unique perspective for learning science that is distinct from most other neuroscience methodologies because it directly reflects neural activities and provides a temporally precise, continuous, and multidimensional view of the cognitive neural processing associated with learning. For example, the unobtrusive and continuous assessment of learners' mental states in real time based on the EEG data opens up the possibility for digital learning platforms to track learners' states and constantly adapt the learning materials to each learner's capacity. Below, we summarize the basic principles of the EEG technique and review how these techniques are advantageous for digital learning science.

EEG is well known for its high temporal resolution, as it measures the instantaneous voltage changes from the scalp with no delay from the actual neural activity in the brain. EEG, typically recorded by electrodes placed on the scalp, mainly reflects the postsynaptic potentials (PSPs) summed from a large population of neurons that are radially oriented near the scalp and activated synchronously. Direct neural activity measures like EEG stand in contrast to the BOLD signals used in fMRI research, reflecting cerebral blood flow subsequent to neural activity, which change too slowly to permit most cognitive processes to be measured in real time. Thus, compared to a questionnaire-based methodology in reflecting learners' mental states where responses tend to reflect only the moments immediately before the questionnaire was taken, an EEG assessment captures changes in real-time mental states of learners (e.g., cognitive load, emotions, fatigue, or motivations), which can then provide immediate feedback or to individualize the learning materials in digital learning environments.

An EEG signal can be decomposed into multiple frequencies through time-frequency analysis, and EEG-based passive brain–computer interfaces (pBCI) which provides efficient real-time quantification of learners' brain activities that was difficult to achieve with manual coding have become an important tool in learning science research. Frequency components of an EEG signal are usually quantified in terms of power (amplitude squared) at each frequency over time. Different frequency bands have been identified and labeled in the literature, including delta (δ: ∼0.2–3.5 Hz), theta (θ: ∼4–7.5 Hz), alpha (α: ∼8–13 Hz), beta (β: ∼14–30 Hz), gamma (γ: ∼30–90 Hz), and (very) high frequencies (> 90 Hz) (Biasiucci, Franceschiello, & Murray, 2019). The frequency bands have significance in various cognitive processes. For example, delta is typically observed when a person is sleeping, theta is typical of nervousness, an attentive and relaxed state of mind is characterized by alpha, alertness is characterized by beta, and problem solving or higher cognitive functions are associated with gamma. One thing to note is that by decomposing the EEG signal into its constituent frequencies, some temporal resolution is sacrificed; as in signal processing, temporal precision is inversely related to frequency precision.

Continuous EEG data can also be used to derive event-related potentials (ERPs). By aligning and averaging point-by-point over multiple segments of EEG data that are time-locked to a particular sensory, cognitive, or motor event of interest, random fluctuations in the EEG signal are cancelled or attenuated, leaving voltage fluctuations that have a systematic temporal relation with event onset. Unlike the frequency approach which inevitably sacrificed some degrees of temporal resolution, ERPs reflect the moment-by-moment fluctuation of brain activity to the millisecond; thereby allowing for the continuous monitoring of processing and the measurement of temporally transient effects. ERP data are also multi-dimensional, with polarity, amplitude, latency, and scalp distribution potentially linked to different aspects of the brain functioning in question. ERP data thus offer the opportunity to tease apart cognitive sub-processes related to learning that are not distinguishable in behavioral measures and/or that may occur too quickly to be captured by most other methods.

ERP data are often discussed in terms of components, which are systematic patterns of voltage changes in magnitude, timing, or scalp region, that can be linked to certain neural and psychological processes and/or certain brain systems (Luck, 2014). The tremendous number of published studies using ERP components, especially in the field of learning and memory, have produced a great deal of knowledge about component properties and characterizations and the factors that may influence the magnitude and timing of these components (for comprehensive reviews on ERP components, see Luck & Kappenman, 2013). This knowledge is thus advantageous and essential for designing digital learning studies, as it provides a solid basis for hypothesis testing and meaningful data interpretation.

1.2. EEG and ERP research in digital learning environments

Below we reviewed some examples of EEG and ERP studies in recent digital learning research, highlighting two major types of applications—EEG studies that assessed student's online mental states to adjust learning environments and ERP studies that assessed brain responses time-locked to specific events to infer learner's progress and identities.

One emphasis of digital learning research is to probe into learners' real-time mental states in a less interruptive manner and use that information to provide more customized and dynamic digital learning environment. For example, cognitive theories of instructional design hold that the type and amount of working-memory load (WML) that learners experience is crucial for successful learning (e.g., Mayer, 2009; Sweller, van Merrienboer, & Paas, 1998). In addition, affect functions have also been thought to play a critical role in learning and learning motivation (Ge, Zhang, Li, & Su, 2019; Keller & Suzuki, 2004; Pentaraki & Burkholder, 2017). Direct and undisruptive evaluation of learner's cognitive load and affective states is therefore essential to provide learning conditions with the optimal level of challenge that can reduce boredom and off-task behavior. These learner states have traditionally been assessed with post hoc questionnaires. Some studies relied on spontaneous biophysical signals of learners, such as facial expression, eye gaze, voice, skin conductance, blood pressure, heart rate, and body language. However, data analysis usually required manual categorization of the states from trained coders which could be time consuming and prone to inter-rater variability.

The EEG technique is ideally suited to help address the above-mentioned difficulties. Continuous whole-head EEG recordings during learning sessions can be classified into different patterns by machine learning algorithms and signal processing techniques for subsequent analysis with a decent level of classification accuracy within or across individuals (Gerjets, Walter, Rosenstiel, Bogdan, & Zander, 2014; Spüler et al., 2016; Wang, Nie, & Lu, 2014). For example, Conrad and Bliemel (2016) found that the average EEG alpha (8–13 Hz) was higher for materials that were appropriately challenging, but EEG beta (13–30 Hz) was lower when the challenge and skill dimensions were low. EEG-assessed cognitive load during learning has also been shown to be indicative of learners' attention and performance (Gaume, Dreyfus, & Vialatte, 2019; Hu, Li, Sun, & Ratcliffe, 2018; Mills et al., 2017) as well as processing demand imposed by the learning materials (e.g., lower cognitive load for processing 3D visuals than 2D visuals in a virtual learning environment, Dan & Reiner, 2017). Furthermore, students' attention and emotion (valence and arousal) can be classified by the fine K-Nearest Neighbor (KNN) algorithm with the EEG features when they are involved in the virtual reality (VR) courses or real lectures (Alwedaie, Khabbaz, Hadi, & Al-Hakim, 2018). These results not only provide objective and immediate assessments for understanding whether students are engaged in the course materials and the style designs, but also offer neurophysiological evidence indicating why virtual learning is more effective than traditional lecturing (Moazami, Bahrampour, Azar, Jahedi, & Moattari, 2014).

EEG-assessed mental state information has also been used to improve learning materials and tailor them to students' needs (Santos, 2016). For example, taking the multi-modal approach of measuring EEG signals in conjunction with blood pressure and skin conductance, Shen and colleagues demonstrated the feasibility of using physiological data to detect learners' real-time emotional states and to feed these states into a digital learning model to automatically adjust the content (e.g., deliver examples or case studies for the current topic when confusion is detected, or deliver music to a student's taste when hopefulness is detected) (Shen, Wang, & Shen, 2009). EEG-assessed cognitive load during learning has been used to provide immediate feedback and to prompt an interactive online learning environment to automatically adjust the difficulty of the learning materials to place them in the optimal range for a particular learner (Mora-Sánchez, Pulini, Gaume, Dreyfus, & Vialatte, 2020; Walter, Rosenstiel, Bogdan, Gerjets, & Spüler, 2017).

In addition to using frequency decomposition to quantify and distinguish different mental states, another main application of EEG technique is to derive ERPs to investigate brain responses that are phase-locked and time-locked to specific events. As the ERP technique requires averaging over multiple second-long EEG observations time-locked to event onsets, research taking this approach is usually performed in a more controlled setting to ensure that the EEG segments averaged together are elicited by similar events. With knowledge of the functional characteristics of many ERP components (Luck & Kappenman, 2013), ERPs have been used to investigate difficulties or learning progress of students (Brown, Howardson, & Fisher, 2016). For example, Conrad and Newman (2019) used the oddball P300 to detect mind-wandering among learners. Huang and Liu (2012) discovered that high- and low-achieving students used different mental rotation strategies (indexed by the ERP rotation-related negativity) while learning chemical structural formulas. Furthermore, Osterhout and colleagues (2006) demonstrated changes in brain responses in L2 learners as their proficiency progressed: while novice learners initially treated morphosyntactic errors as lexical or semantic errors (as indexed by an N400 effect), these learners' brain responses changed over the course of learning to approximate brain responses to morphosyntactic errors in their native language (as indexed by a P600 effect).

As some ERP components can be used to represent personal idiosyncratic processing, the ERP technique has also been applied to provide reliable continuous authentication to ensure that the identity of the individual does not change after logging in. Prior research has shown that some ERP components have robust identifiable features that can be used to differentiate the brain responses of different individuals. For example, the N400 response is thought to reflect idiosyncratic semantic experiences (Coronel & Federmeier, 2016), and the P300 has been associated with attention-mediated processes that vary across individuals (Polich, 2012). Taking advantage of these component characteristics, Song and colleagues proposed that P300 in conjunction with eye tracking could be used as biometrics for continuous personal identification (Song, Lee, & Nam, 2013). Applying pattern classifiers on ERP responses to a stream of text designed to be idiosyncratically familiar to different individuals, Armstrong and colleagues also demonstrated decent accuracy in identifying the individuals responsible for particular ERP responses (Armstrong et al., 2015).

2. Experimental design suggestions

Applying measurements of brain responses to the learning science field has the potential to provide input into the development of learning materials and facilitate the design of individualized learning environments. However, this type of research involves digital learning system development, experimental design, and neurophysiological signal processing, and is thus highly interdisciplinary. Additionally, EEG and ERP data are complex and multifaceted, which is an advantage of the technique but also makes data analysis challenging. A further challenge is that EEG and ERP data are sensitive to the various types of stimuli, the age range of participants, and uncontrolled experimental factors. Thus, collaboration with experts to identify a suitable experimental design and potential confounding factors is advised to ensure better research quality and more accurate data interpretations when studying EEG/ERP responses. Some suggestions for conducting this type of research are provided below.

To successfully integrate and apply neurophysiological results, researchers must first state the research question concretely and provide clear operational definitions of the cognitive processes and/or mental states to be examined. For example, when examining an overarching concept of engagement during a task, this could be defined as focused attention on the task or as a high cognitive load being required for the task (Brouwer, Zander, van Erp, Korteling, & Bronkhorst, 2015). It is possible that researchers using the same term may indicate a different underlying cognition across research fields, and thus it is important to reduce confusion by providing clear definitions. Second, researchers should formulate a linking hypothesis of which EEG/ERP measures are expected to vary with the cognitive processes/mental states in the study design. Drawing conclusions about cognition from EEG/ERP data requires inferences, and it is therefore important to know what assumptions are being made to permit these inferences. In other words, a detailed literature review is necessary to establish what independent variables do and do not influence the EEG/ERP measures before conducting the research (Cacioppo & Tassinary 1990). All inferences are correlational in nature—even when a tight correlation is found indicating that the linking hypothesis is correct, only a correlate of cognitive processes/mental states has been discovered. The measured brain activity cannot be interpreted as the direct manifestation of the cognitive processes/mental states that it has been linked to (Handy, 2005). Third, researchers need to be aware of confounding factors. EEG/ERP data are very sensitive to various types of factors that may not be controlled for in the study design, such as the modality of the stimulus presentation (visual versus auditory), the characteristics of the stimuli, such as familiarity with the materials (Kutas & Federmeier, 2011) and the concreteness (Huang, Lee & Federmeier, 2010; Huang & Federmeier, 2015) or ambiguity of the words (Huang & Lee, 2018), and repetitions (Kutas & Federmeier, 2011). The syntactic word ordering effect may actually reflect the uncontrolled concreteness differences between words (Huang & Federmeier, 2012). Fourth, because only EEG/ERP data that are associated with the cognitive processes of interest and not contaminated by artifacts will be included in the data analysis, researchers should identify tasks that require minimal gross motor movement. Fifth, researchers must assess and evaluate whether their given set of assumptions is warranted based on how the brain signals are collected and analyzed.

In Figure 1, we summarize some important experimental design suggestions. At the beginning of conducting a research, investigators have to consider which EEG system would be more suitable for addressing their research questions, this issue will be elaborated in the next session again. In some of the digital learning studies, the different teaching methods (e.g., digital learning vs. lecture-based training) are composed of different groups of participants; in this case researchers have to make sure the characteristics of participants in different groups are matched (such as age range, gender, participants' intelligence). Another issue that needs to be taken into account in the experimental design is the total duration of a study, since the EEG signal quality would be substantially reduced when participants are tired. It would be ideal to split one study into multiple short blocks with about ten to fifteen minutes per block. Between blocks, the participants could take a short break. When having multiple blocks in a study, then it is not good to administer the blocks to all of the participants in the same sequence. That means, a randomized counterbalancing is suggested to reduce the order and carryover effects. On the other hand, if the duration of the whole study is longer than one hour, then the researchers could seek the possibility of asking participants to join a multi-session study. Otherwise, splitting one study into two shorter studies could be another option. As for how many participants are needed to test the hypothesis adequately, the decision is based on the study design, the variability of the data and the type of statistical procedure to be used.

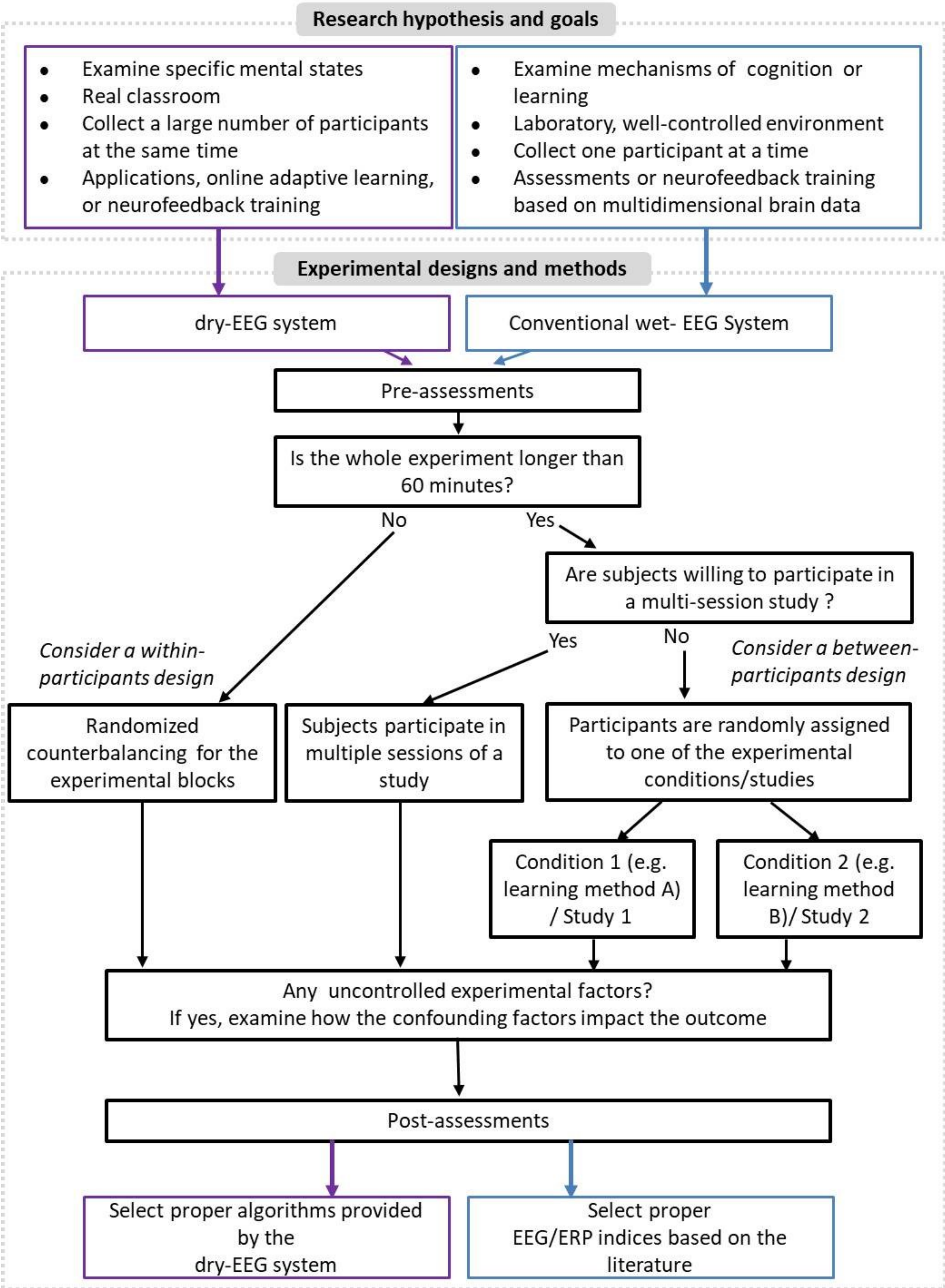


Figure 1. Summary of the experimental design suggestions

3. How to record publishable data from the lab to the field

In this session, we present major EEG/ERP data collection techniques and preprocessing methods that can be used beyond laboratory settings. Some suggestions are also offered for planning studies toward real-world measurements.

3.1. EEG data recording in the laboratory

Brain signals are small along the scalp surface, in the order of 10–100 microvolts, whereas non-EEG biological signals, such as skin potentials, muscle activities, blinks, and eye-movements, are in the order of 50–500 microvolts. To better read the brain signals, they need to be amplified and filtered by the recording system. When recording EEG signals, the most often used filtering ranges are within 0.01 to 30 Hz or, more conservatively, 0.01 to 100 Hz. It is important to keep most of the low frequency activity, but very high frequency activity can be safely discarded because it is unlikely to be biologically related. The continuous analog signal emitted from the amplifier must then be converted to a digital format (through A/D conversion), forming a discrete pairing of time points and voltages. Once the signal is represented as individual numbers in a time series, these numbers can be manipulated mathematically. Based on the Nyquist criterion, the minimum sampling rate recommended is at least twice as fast as the fastest frequency component in the signal. For example, to investigate an EEG signal of 40Hz, the sampling rate must be at least 80Hz. However, most researchers sample at four to eight times the highest frequency to ensure accurate detection of the EEG data and/or under the consideration that greater temporal resolution might be needed. Finally, EEG activity from the various sites on the scalp needs to be referenced to the mastoid(s) or earlobe(s), common locations that pick up minimal amounts of brain activity, to give the difference between each site and the reference electrode.

A system for presenting learning materials, pre- and/or post-testing, and receiving behavioral responses needs to communicate with the EEG data acquisition system in real time to send event codes whenever an event occurs (e.g., a stimulus being presented or a response generated by the subject). These event codes are used as the time-locking points for data processing or signal averaging, so the timing must be precise: the EEG acquisition system will not “know” what kind of event the subject is encountering nor at what time if no event codes are given.

Before data collection, researchers need to ensure that the areas under the EEG electrodes are free of dead skin, oil, and sweat, for low impedance. High electrode impedance increases noise and decreases statistical power, meaning that more trials are needed to reach statistical significance. Typically, impedance under 5 kΩ is suggested with the use of conductive gel (so called wet-EEG). The electrode positions should follow the standard International 10–20 system, because the relations between the 10–20 electrode system and the underlying cortical anatomy have been validated (Towle et al., 1993). A comfortable environment is also essential, and the experimental room should have an air conditioning system that can control the temperature, humidity, and air circulation for indoor air quality. Because sweat is one of the common causes of biological artifacts that can alter impedance, and high temperature also causes participants to become sleepier. Additionally, to minimize any muscle activities or body movements during the EEG/ ERP recording sessions, it is very important for the subject to be seated in a comfortable position.

3.2. EEG data recording in real-world settings

Compared with collecting EEG data in the laboratory, where recordings can be better controlled and monitored by the experimenters, collecting EEG data in the “real world” poses additional challenges. For example, gel application and post-recording cleanings are time-consuming. For real-world applications, dry-wireless EEG systems (dwEEGs), which are wireless systems using dry electrodes to collect EEG signals, have been developed over the past decade (Di Flumeri et al., 2019). Compared with the conventional wet-EEG devices, dwEEGs can shorten the time for preparation, remove the need to apply conductive gel, increase the convenience to new users, and enlarge the number of simultaneous participants for investigating their interactions. In Table 1, we contrasted features between wet-EEG and dry-EEG for studying digital learning topics. Although some limitations (e.g., the comfort of the sensor) are still being addressed (Di Flumeri et al., 2019; Lin, Yu, King, Liu, & Liao, 2020), dwEEGs already demonstrate their usefulness in many areas, including brain–computer interface (BCI) (Lin et al., 2018), sport science (Wang, Moreau, & Kao, 2019), clinical assessment (Lin et al., 2017; Ratti et al., 2017), and education (Xu & Zhong, 2018).

Table 1. Comparisons between the wet-EEG and dry-EEG devices for a digital learning research.

Device	wet-EEG	dry-EEG
Cost	High (USD 40,000 – 80,000)	Low (USD 100 – 15,600)
Channel number	32 to 256	1 to 64
Preparation	Experience required	Easy
Setting time	15–30 minutes	1–10 minutes
Wash hair	Yes	No
Comfortability	Average	Comfort to average
Signal stability	Very High	Very low to high
Weight	average	Depends on the channel number
Wireless	Yes	Yes
Scalable	Flexible	Fixed
Raw data	Available	Less available to available
Algorithm	Less available	Available
Artifact removal	Available	Less available to available
Analysis	All analysis, including brain connection, source, time and frequency domains	Time and frequency domain
Subjects at the same time	Most are one	Not limited
Research Topics	All topics and applications	Many are for applications (monitoring attention, emotion or meditation states)
Summary	Convectional and reliable tool for educational research, especially for collecting a single subject data at a time in the laboratory.	Convenient device to explore educational topics in the real environment, especially for evaluating specific processing/mental states of a larger population at the same time.

Some types of dwEEG (see Figure 2) have been used in real classroom environments to examine the relationships between dynamic changes in the brain and learning effects (Lau-Zhu, Lau, & McLoughlin, 2019). Although the first two systems (a and b) are designed for ordinary consumers and the others (c-f) are more for scientific research, all of these systems have been used in educational studies. Generally, these systems consist of a miniaturized light-weight amplifier with a Bluetooth module to transmit the signals, which increases the wireless and portability of each device (Bateson, Baseler, Paulson, Ahmed, & Asghar, 2017). The devices are light (under 269 grams), the sampling rate is reasonable (above 256 Hz), and the battery life of most dwEEGs is good (above 5 hours), supporting the criteria for standard EEG settings for regular experimental sessions. Each system has its own software to support the EEG signal collection, and most can save the raw data in text format for further signal processing using other software, such as EEGLAB (Delorme & Makeig, 2004). Most importantly, the signal quality and comfort of these systems have been validated following the traditional EEG paradigm with wet-EEG devices in the frequency domain (i.e., typically between 1–50 Hz, e.g., alpha band) and time domain (i.e., identifying time-locked brain electrical activity to a stimulus, e.g., P300) (Ruffini et al., 2006; Badcock et al., 2015; Oliveira, Schlink, Hairston, König, & Ferris, 2016; Williams, Norton, Hassall, & Colino, 2017; Rieiro et al., 2019; Kam et al, 2019; Lin et al., 2020). Although dwEEGs have been shown to benefit education studies (Xu & Zhong, 2018), the diversity of manufacturers still creates challenges for researchers in deciding which system to use and how many electrodes/channels are required for a study. Three challenges are described below for researchers to consider when planning their studies.

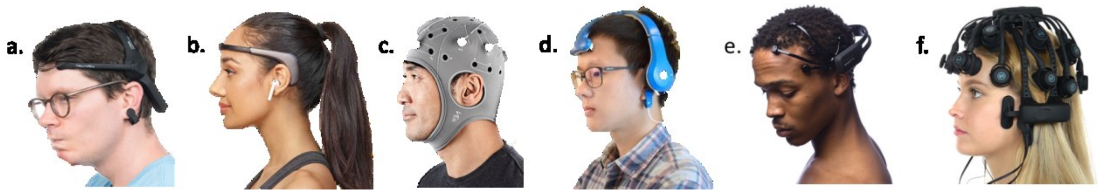


Figure 2. Examples of dry-wireless EEG systems. (a) MindWave 2 (NeuroSky); (b) Muse (InteraXon Inc.); (c) ENOBIO 8 (Neuroelectrics); (d) BR8 (BRI); (e) EPOC (Emotiv); (f) Quick (Cognionics)

First, measuring errors can occur due to low numbers of channels. Although fewer channels can reduce the cost and could be more comfortable for the participants due to a lighter weight, it will not only dramatically increase the application limitations but also the measuring errors (Xu & Zhong, 2018). For example, the frontal lobe is thought to be associated with emotions and motivation (Pessoa, 2009). However, this assumption is based on the asymmetry between two hemispheres and is most commonly computed by subtracting the natural log of the left side's alpha power (often using electrode F3) from the natural log of the right side's alpha power (electrode F4) (Coan & Allen, 2004). A single channel (e.g., the FP1 of NeuroSky) is not able to indicate frontal asymmetry--an effective index for emotion identification.

Second, measuring limitations can arise due to the layout of the channels. The final optimal design of dwEEGs is to cover the standard areas following the International 10–20 system used in the conventional wet-EEG. However, the layout of some dwEEGs is limited by the characteristics of the electrodes and cannot fit the 10–20 system. For example, although Emotiv possess 14 electrodes (wet), it does not cover the midline of the scalp, especially the central and parietal areas. These sites often evoke some unique components of ERPs that are associated with the motor and attention functions. For instance, a well-known P300 (P3b) is elicited over the parietal site (Pz electrode) when stimulus detection engages memory operation (Polich, 2012). The P300 latency and amplitude follow a maturational path from childhood to adolescence (van Dinteren, Arns, Jongsma, & Kessels, 2014) and are treated as an important index to investigate typical and atypical development. A lack of information on the parietal area places limitations on the use of P300 as an informative index for a study.

Third, the provided algorithms are short on validation. Some dwEEGs provide users with their own algorithms for monitoring the subject's attention level, emotions, and mediation states (e.g., Neurosky). Although most algorithms have corresponding technical support, the manufacturers do not provide clear statements on how the data are calculated. Moreover, the channel layout and channel numbers are still core problems for developing algorithms. Even for the few algorithms with high classification accuracy, cumulative samples are not enough to extend the usage range, especially for children (Xu & Zhong, 2018). Because developmental differences can influence the engagement of neural oscillations (Schneider, Abel, Ogiela, McCord, & Maguire, 2018), more evidence is required to validate the effectiveness of these algorithms before applying them in a real educational environment.

In sum, dwEEGs could provide good opportunities to study education/digital learning through measurements of brain activities. However, researchers need to consider the measuring limitations based on their research goal and budget before purchasing these devices. Researchers who hope to turn their research findings into practical achievements should be cautious in adopting the user-friendly interfaces and convenient algorithms provided by some dwEEGs. Three suggestions can be considered before deciding on the system of dwEEGs. First, if your study is an exploratory research, dwEEGs with more channels would be preferred, as more channels can cover more of the important scalp areas that provide critical information for further and diversity of analysis. Second, if you intend to use the same indexes as previous studies did, then exact corresponding channels should be involved in the layout of the selected dwEEGs. Third, if you focus on a particular application (e.g., attention monitoring of the digital learning process), you can use the algorithms provided by the device as they have been validated and published in peer-review journals with similar experiment designs and participants as yours.

3.3. Data preprocessing and writing-up

After data collection is completed, it is essential to remove the EEG signals that are contaminated by artifacts (both non-EEG biological and environmental electrical noises) prior to data analysis, as such artifacts may lead to misinterpretation of the EEG/ERP results. Eye movements and blinks can easily be identified on the electrodes attached to the supra-outer canthus of the left eye and infra-outer canthus of the right eye or on the frontal electrodes. A variety of techniques are available for removing these artifacts (see review in Urigüen, & Garcia-Zapirain, 2015; Jiang, Bian, & Tian, 2019) and can be primarily categorized into two approaches: estimation by reference channels and decomposing signals into different domains. The typical method of the first approach is the regression analysis which calculates the amplitude difference between reference channels (usually EOG/ECG) and other EEG channels, and then subtracts the estimated artifacts from EEG (Sweeney, Ward, & McLoone, 2012). This type of method is usually applied to remove eye movements. The second approach involves the independent component analysis (ICA) that can be used to remove all kinds of artifacts. This technique assumes that the EEG signal is a linear mixture of brain signals and artifacts which can be decomposed into independent components (ICs) (Makeig, Bell, Jung, & Sejnowski, 1996) to discard the artifacts and reconstruct the clean signals for further analysis. Recently, many researchers have developed different hybrid models to improve the accuracy and efficacy of the automatically processing (e.g., Icaeyeblinkmetrics toolbox; Pontifex, Miskovic, & Laszlo, 2017; artifact subspace reconstruction method; Chang, Hsu, Pion-Tonachini, & Jung, 2018). However, these must be applied with caution to ensure they do not distort the brain signals. Once the EEG signals are noise-free, further data analyses can be conducted.

The development of learning science depends on the validation of the results. Researchers therefore need to ensure that the details of the study are documented sufficiently so that others can evaluate and replicate published results. Specifically, other than the rationale for the proposed study and the discussion of the results, some details must be reported in writing up a paper with EEG/ERP results. First, detailed and clear descriptions are needed of the EEG measurements, such as the acquisition protocol, EEG data acquisition system, locations of the recording electrodes on the scalp, reference electrode(s), and methods and criteria that were used to remove the artifacts. Second, basic characteristics of the participants, such as mean age, age range, gender, and handedness must be reported. Third, the details of the materials used in the study and how the materials were presented to the subjects also need to be described. And finally, proper graphical representation of the EEG/ERP results on the paper is necessary. All these aspects are essential in a manuscript as suggested by the Society for Psychophysiological Research committee's guidelines (Keil et al., 2014; Picton et al., 2000). These guidelines not only provide a solid foundation for using or applying EEG technology in understanding human cognition but are also very helpful in communicating or comparing EEG/ERP results between studies.

4. Potential research directions and concluding remarks

We highlight three potential areas for new research. The first is to investigate interactions in the classroom setting. Researchers can use EEG as a more objective and naturalistic approach to investigate the effect of teaching materials on students or the interactions between a teacher and students or between students. For example, Dikker et al. (2017) used Emotiv to simultaneously record the brain activities of 12 high school students to identify neural markers of group engagement in a real-world classroom. Using brain-to-brain synchrony (BBS) as a neural marker, students with higher BBS showed greater engagement in the group. Moreover, students who showed higher BBS with the teacher reported greater social closeness to the teacher, especially when the teacher was giving a lecture (Bevilacqua et al., 2019). These findings suggest that BBS may reflect social interaction in real-world group settings and show dwEEGs to be useful tools for investigating the social relationships and interactions taking place in a real classroom setting.

The second research suggestion is to explore specific experiences with large-scale data collection. Educators are among those needing a way to naturalistically evaluate the specific experience elicited by a work of art, design product, or set of learning materials. Using a portable EEG device, a 400-person EEG dataset was collected to examine the neural basis of aesthetic experiences during visual exhibitions (Kontson et al., 2015). The connection strength in localized brain networks was significantly increased while subjects viewed the most aesthetically pleasing art compared to a blank wall. Moreover, the direction of EEG signal flow showed the early recruitment of broad posterior areas followed by focal anterior activation. This example shows that dwEEGs may be helpful in understanding how the brain integrates sensory input and ongoing internal states to produce specific phenomena, such as the aesthetic experience.

A third area of research potential is maximizing the efficiency of adaptive/online learning by real-time neuro-feedback. At the time of writing, the need for online courses has dramatically increased due to the COVID-19 pandemic. However, instructors are unable to immediately monitor whether students remain focused on their learning. With one advantage of dwEEGs being their capacity to provide convenient and real-time measures for monitoring participants' attention and emotional states, these devices can be used as a novel way to monitor online learning processes and elevate the efficiency of e-learning environments. To examine students' attention levels during digital learning, the Neurosky team has developed an attention aware system integrated with a video lecture tagging system (Chen, Wang, & Yu, 2017). Negative correlations were found between learning performance and low levels of attention to video lectures. In another study using Emotiv, some important features associated with the learning performance of Vietnamese language via video were extracted and used as indices to adjust the instructional methods and/or materials for adaptive learning (Hu & Kuo, 2017). Therefore, dwEEGs may provide a convenient approach to enhance the efficiency of online and adaptive learning for instructors or users within the digital learning environment.

Over the past half-century, EEG and ERP studies have opened up new avenues for understanding human cognitive processes as they occur in real time. Applying these techniques with competence, caution, and creativity can aid in the development of productive learning environments. Such advancements in learning science can ensure that everybody can learn effectively, both inside and outside of school.

Acknowledgement

We thank Prof. Kara Federmeier for her constructive comments. This work was supported by City University of Hong Kong SRG-Fd 7005343, and by Hong Kong Institute for Advanced Study (9360157).

References

Alwedaie, S. A., Khabbaz, H. A., Hadi, S. R., & Al-Hakim, R. (2018). EEG-based analysis for learning through virtual reality environment. *Journal of Biosensors & Bioelectronics, 9*, 1-6. doi:10.4172/2155-6210.1000249

Arguel, A., Lockyer, L., Lipp, O. V., Lodge, J. M., & Kennedy, G. (2017). Inside out: Detecting learners' confusion to improve interactive digital learning environments. *Journal of Educational Computing Research, 55*(4), 526–551. doi:10.1177/0735633116674732

Armstrong, B. C., Ruiz-Blondet, M. V., Khalifian, N., Kurtz, K. J., Jin, Z., & Laszlo, S. (2015). Brainprint: Assessing the uniqueness, collectability, and permanence of a novel method for ERP biometrics. *Neurocomputing, 166*, 59–67. doi:10.1016/j.neucom.2015.04.025

Badcock, N. A., Preece, K. A., de Wit, B., Glenn, K., Fieder, N., Thie, J., & McArthur, G. (2015). Validation of the Emotiv EPOC EEG system for research quality auditory event-related potentials in children. *PeerJ, 3*, e907. doi:10.7717/peerj.907

Bateson, A. D., Baseler, H. A., Paulson, K. S., Ahmed, F., & Asghar, A. U. R. (2017). Categorisation of mobile EEG: A Researcher's perspective. *BioMed Research International, 2017*, 5496196. doi:10.1155/2017/5496196

Bevilacqua, D., Davidesco, I., Wan, L., Chaloner, K., Rowland, J., Ding, M., Poeppel, D., & Dikker, S. (2019). Brain-to-brain synchrony and learning outcomes vary by student-teacher dynamics: Evidence from a real-world classroom electroencephalography study. *Journal of Cognitive Neuroscience, 31*(3), 401–411. doi:10.1162/jocn_a_01274

Biasiucci, A., Franceschiello, B., & Murray, M. M. (2019). Electroencephalography. *Current Biology, 29*(3), R80–R85. doi:10.1016/j.cub.2018.11.052

Brouwer, A.-M., Zander, T. O., van Erp, J. B. F., Korteling, J. E., & Bronkhorst, A. W. (2015). Using neurophysiological signals that reflect cognitive or affective state: Six recommendations to avoid common pitfalls. *Frontiers in Neuroscience, 9*, 136. doi:10.3389/fnins.2015.00136

Brown, K. G., Howardson, G., & Fisher, S. L. (2016). Learner control and e-learning: Taking stock and moving forward. *Annual Review of Organizational Psychology and Organizational Behavior, 3*(1), 267–291. doi:10.1146/annurev-orgpsych-041015-062344

Cacioppo, J. T., & Tassinary, L. G. (1990). Inferring psychological significance from physiological signals. *American Psychologist, 45*(1), 16–28. doi:10.1037/0003-066X.45.1.16

Chang, C. Y., Hsu, S. H., Pion-Tonachini, L., & Jung, T. P. (2018). Evaluation of artifact subspace reconstruction for automatic EEG artifact removal. In *Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1242-1245). doi:10.1109/EMBC.2018.8512547

Chen, C. M., Wang, J. Y., & Yu, C. M. (2017). Assessing the attention levels of students by using a novel attention aware system based on brainwave signals. *British Journal of Educational Technology, 48*(2), 348–369. doi:10.1111/bjet.12359

Coan, J. A., & Allen, J. J. (2004). Frontal EEG asymmetry as a moderator and mediator of emotion. *Biological Psychology, 67*(1–2), 7–49. doi:10.1016/j.biopsycho.2004.03.002

Conrad, C., & Bliemel, M. (2016, December). *Psychophysiological measures of cognitive absorption and cognitive load in e-learning applications.* Paper presented at the International Conference on Information Systems (ICIS), Dublin, Ireland.

Conrad, C., & Newman, A. (2019). Measuring the impact of mind wandering in real time using an auditory evoked potential. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, & A. B. Randolph (Eds.), *Information systems and neuroscience* (pp. 37–45). Cham, Switzerland: Springer International Publishing. doi:10.1007/978-3-030-01087-4_5

Coronel, J. C., & Federmeier, K. D. (2016). The N400 reveals how personal semantics is processed: Insights into the nature and organization of self-knowledge. *Neuropsychologia*, *84*, 36-43. doi:10.1016/j.neuropsychologia.2016.01.029

Dan, A., & Reiner, M. (2017). EEG-based cognitive load of processing events in 3D virtual worlds is lower than processing events in 2D displays. *International Journal of Psychophysiology, 122*, 75–84. doi:10.1016/j.ijpsycho.2016.08.013

Delorme, A., & Makeig, S. (2004). EEGLAB: An Open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21. doi:10.1016/j.jneumeth.2003.10.009

Di Flumeri, G., Aricò, P., Borghini, G., Sciaraffa, N., Di Florio, A., & Babiloni, F. (2019). The Dry revolution: Evaluation of three different EEG dry electrode types in terms of signal spectral features, mental states classification and usability. *Sensors, 19*(6), 1365. doi:10.3390/s19061365

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., Rowland, J., Michalareas, G., Van Bavel, J. J., Ding, M., & Poeppel, D. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current Biology, 27*(9), 1375–1380. doi:10.1016/j.cub.2017.04.002

Fabiani, M., Gratton, G., & Federmeier, K. D. (2007). Event-related brain potentials: Methods, theory, and application. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 85-119). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511546396.004

Fischer, F., Hmelo-Silver, C. E., Goldman, S. R., & Reimann, P. (2018). *International handbook of the learning sciences*. New York, NY: Routledge.

Gaume, A., Dreyfus, G., & Vialatte, F. B. (2019). A Cognitive brain–computer interface monitoring sustained attentional variations during a continuous task. *Cognitive Neurodynamics, 13*(3), 257–269. doi:10.1007/s11571-019-09521-4

Ge, Z., Zhang, A., Li, Y., & Su, J. (2019). Exploring the impact of teachers' verbal immediacy as an emotion mediating factor on adult e-learners' language learning. *Educational Technology & Society, 22*(4), 77-89.

Gerjets, P., Walter, C., Rosenstiel, W., Bogdan, M., & Zander, T. O. (2014). Cognitive state monitoring and the design of adaptive instruction in digital environments: Lessons learned from cognitive workload assessment using a passive brain-computer interface approach. *Frontiers in Neuroscience, 8*, 385. doi:10.3389/fnins.2014.00385

Handy, T. C. (2004). *Event-related potentials: A Methods handbook*. Cambridge, MA: MIT Press.

Hu, B., Li, X., Sun, S., & Ratcliffe, M. (2018). Attention recognition in EEG-based affective learning research using CFS+KNN algorithm. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 15*(1), 38–45. doi:10.1109/TCBB.2016.2616395

Hu, P. C., & Kuo, P. C. (2017). Adaptive learning system for E-learning based on EEG brain signals. In *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)* (pp. 1-2). doi:10.1109/GCCE.2017.8229382

Huang, C.-F., & Liu, C.-J. (2012). An Event-related potentials study of mental rotation in identifying chemical structural formulas. *European Journal of Educational Research, 1*(1), 37–54. doi:10.12973/eu-jer.1.1.37

Huang, H.-W., & Federmeier, K. D. (2012). Dispreferred adjective orders elicit brain responses associated with lexico-semantic rather than syntactic processing. *Brain Research, 1475*, 62–70. doi:10.1016/j.brainres.2012.07.050

Huang, H.-W., & Federmeier, K. D. (2015). Imaginative language: What event-related potentials have revealed about the nature and source of concreteness effects. *Language and Linguistics, 16*(4), 503–515. doi:10.1177/1606822X15583233

Huang, H.-W., Lee, C.-L., & Federmeier, K. D. (2010). Imagine that! ERPs provide evidence for distinct hemispheric contributions to the processing of concrete and abstract concepts. *NeuroImage, 49*(1), 1116–1123. doi:10.1016/j.neuroimage.2009.07.031

Huang, H.-W., & Lee, C.-Y. (2018). Number of meanings and number of senses: An ERP study of sublexical ambiguities in reading Chinese disyllabic compounds. *Frontiers in Psychology, 9*, 324. doi:10.3389/fpsyg.2018.00324

Jiang, X., Bian, G. B., & Tian, Z. (2019). Removal of artifacts from EEG signals: A Review. *Sensors, 19,* 987. doi:10.3390/s19050987

Kam, J. W. Y., Griffin, S., Shen, A., Patel, S., Hinrichs, H., Heinze, H. J., Deouell, L. Y., & Knight, R. T. (2019). Systematic comparison between a wireless EEG system with dry electrodes and a wired EEG system with wet electrodes. *NeuroImage, 184*, 119–129. doi:10.1016/j.neuroimage.2018.09.012

Keil, A., Debener, S., Gratton, G., Junghöfer, M., Kappenman, E. S., Luck, S. J., Luu, P., Miller, G. A., & Yee, C. M. (2014). Committee report: Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology, 51*(1), 1–21. doi:10.1111/psyp.12147

Keller, J., & Suzuki, K. (2004). Learner motivation and E-learning design: A Multinationally validated process. *Journal of Educational Media, 29*(3), 229–239. doi: 10.1080/1358165042000283084

Kontson, K. L., Megjhani, M., Brantley, J. A., Cruz-Garza, J. G., Nakagome, S., Robleto, D., White, M., Civillico, E., & Contreras-Vidal, J. L. (2015). Your brain on art: Emergent cortical dynamics during aesthetic experiences. *Frontiers in Human Neuroscience, 9*, 626. doi:10.3389/fnhum.2015.00626

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621–647. doi:10.1146/annurev.psych.093008.131123

Lan, Y. J. (2020). Immersion into virtual reality for language learning. In K. D. Federmeier & H.-W. Huang (Eds.), *Psychology of learning and motivation: Adult and second language learning* (vol. 72, pp. 1–26). San Diego, CA: Elsevier Science. doi:10.1016/bs.plm.2020.03.001

Lau-Zhu, A., Lau, M. P. H., & McLoughlin, G. (2019). Mobile EEG in research on neurodevelopmental disorders: Opportunities and challenges. *Developmental Cognitive Neuroscience, 36*, 100635. doi:10.1016/j.dcn.2019.100635

Lin, C. T., Chiu, C. Y., Singh, A. K., King, J. T., Ko, L. W., Lu, Y. C., & Wang, Y. K. (2018). A Wireless multifunctional ssvep-based brain–computer interface assistive system. *IEEE Transactions on Cognitive and Developmental Systems, 11*(3), 375–383. doi:10.1109/TCDS.2018.2820153

Lin, C. T., Chuang, C. H., Cao, Z., Singh, A. K., Hung, C. S., Yu, Y. H., Nascimben, M., Liu, Y. T., King, J. T., & Su, T. P. (2017). Forehead EEG in support of future feasible personal healthcare solutions: Sleep management, headache prevention, and depression treatment. *IEEE Access, 5*, 10612–10621. doi:10.1109/ACCESS.2017.2675884

Lin, C. T., Yu, Y. H., King, J. T., Liu, C. H., & Liao, L. D. (2020). Augmented wire-embedded silicon-based dry-contact sensors for electroencephalography signal measurements. *IEEE Sensors Journal, 20*(7), 3831–3837. doi:10.1109/JSEN.2019.2959619

Luck, S. J. (2014). *An Introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Luck, S. J., & Kappenman, E. S. (2013). ERP components and selective attention. In S. J. Luck (Ed.), *The Oxford handbook of event-related potential components* (pp. 295–327). New York, NY: Oxford University Press. doi:10.1093/oxfordhb/9780195374148.013.0144

Makeig, S., Bell, A. J., Jung, T. P., & Sejnowski, T. (1996). Independent component analysis of electroencephalographic data. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), *Advances in Neural Information Processing Systems* (Vol. 8, pp. 145-151). Cambridge, MA: MIT Press.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press.

Mills, C., Fridman, I., Soussou, W., Waghray, D., Olney, A. M., & D'Mello, S. K. (2017). Put your thinking cap on: Detecting cognitive load using EEG during learning. In *Proceedings of the seventh International Learning Analytics & Knowledge Conference* (pp. 80–89). doi:10.1145/3027385.3027431

Moazami, F., Bahrampour, E., Azar, M. R., Jahedi, F., & Moattari, M. (2014). Comparing two methods of education (virtual versus traditional) on learning of Iranian dental students: A Post-test only design study. *BMC Medical Education, 14*, 45. doi:10.1186/1472-6920-14-45

Mora-Sánchez, A., Pulini, A.-A., Gaume, A., Dreyfus, G., & Vialatte, F.-B. (2020). A Brain–computer interface for the continuous, real-time monitoring of working memory load in real-world environments. *Cognitive Neurodynamics, 14*(3), 301–321. doi:10.1007/s11571-020-09573-x

Oliveira, A. S., Schlink, B. R., Hairston, W. D., König, P., & Ferris, D. P. (2016). Proposing metrics for benchmarking novel EEG technologies towards real-world measurements. *Frontiers in Human Neuroscience, 10*, 188. doi:10.3389/fnhum.2016.00188

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A Means for exploring the neurocognition of second language processing. *Language Learning, 56*(s1), 199–230. doi:10.1111/j.1467-9922.2006.00361.x

Pentaraki, A., & Burkholder, G. J. (2017). Emerging evidence regarding the roles of emotional, behavioural, and cognitive aspects of student engagement in the online classroom. *European Journal of Open, Distance and E-Learning, 20*(1), 1–21. doi:10.1515/eurodl-2017-0001

Pessoa, L. (2009). How do emotion and motivation direct executive control? *Trends in Cognitive Sciences, 13*(4), 160–166. doi:10.1016/j.tics.2009.01.006

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., Jr., Miller, G. A., Ritter, W., Ruchkin, D. S., Rugg, M. D., & Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology, 37*(2), 127–152. doi:10.1017/S0048577200000305

Polich, J. (2012). Neuropsychology of P300. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 159–188). New York, NY: Oxford University Press.

Pontifex, M. B., Miskovic, V., & Laszlo, S. (2017). Evaluating the efficacy of fully automated approaches for the selection of eyeblink ICA components. *Psychophysiology, 54*(5), 780-791. doi:10.1111/psyp.12827

Ratti, E., Waninger, S., Berka, C., Ruffini, G., & Verma, A. (2017). Comparison of medical and consumer wireless EEG systems for use in clinical trials. *Frontiers in Human Neuroscience, 11*, 398. doi:10.3389/fnhum.2017.00398

Rieiro, H., Diaz-Piedra, C., Morales, J. M., Catena, A., Romero, S., Roca-Gonzalez, J., Fuentes, L. J., & Di Stasi, L. L. (2019). Validation of electroencephalographic recordings obtained with a consumer-grade, single dry electrode, low-cost device: A Comparative study. *Sensors (Basel), 19*(12). doi:10.3390/s19122808

Ruffini, G., Dunne, S., Farrés, E., Watts, P. C., Mendoza, E., Silva, S. R., Grau, C., Marco-Pallarés, J., Fuentemilla, L., & Vandecasteele, B. J. (2006). ENOBIO—first tests of a dry electrophysiology electrode using carbon nanotubes. In *Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 1826-1829). doi:10.1109/IEMBS.2006.259248.

Santos, O. C. (2016). Emotions and personality in adaptive e-learning systems: An Affective computing perspective. In M. Tkalčič, B. De Carolis, M. de Gemmis, A. Odić, & A. Košir (Eds.), *Emotions and personality in personalized services: Models, evaluation and applications* (pp. 263–285). Cham, Switzerland: Springer International Publishing. doi:10.1007/978-3-319-31413-6_13

Schneider, J. M., Abel, A. D., Ogiela, D. A., McCord, C., & Maguire, M. J. (2018). Developmental differences in the neural oscillations underlying auditory sentence processing in children and adults. *Brain and Language, 186*, 17–25. doi:10.1016/j.bandl.2018.09.002

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

Song, K.-S., Lee, S. M., & Nam, S. (2013). Cognitive biometrics application for e-learning security enhancement. *International Journal of Bio-Science and Bio-Technology, 5*(3), 143–152.

Spüler, M., Walter, C., Rosenstiel, W., Gerjets, P., Moeller, K., & Klein, E. (2016). EEG-based prediction of cognitive workload induced by arithmetic: A Step towards online adaptation in numerical learning. *ZDM Mathematics Education, 48*(3), 267–278. doi:10.1007/s11858-015-0754-8

Sweeney, K. T., Ward, T. E., & McLoone, S. F. (2012). Artifact removal in physiological signals practices and possibilities. *IEEE Transactions on Information Technology in Biomedicine, 16,* 488–500. doi:10.1109/TITB.2012.2188536

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296. doi:10.1023/a:1022193728205

Towle, V. L., Bolaños, J., Suarez, D., Tan, K., Grzeszczuk, R., Levin, D. N., Cakmur, R., Frank, S. A., & Spire, J.-P. (1993). The Spatial location of EEG electrodes: Locating the best-fitting sphere relative to cortical anatomy. *Electroencephalography and Clinical Neurophysiology, 86*(1), 1–6. doi:10.1016/0013-4694(93)90061-Y

Urigüen, J. A., & Garcia-Zapirain, B. (2015). EEG artifact removal-state-of-the-art and guidelines. *Journal of Neural Engineering, 12*(3), 031001. doi:10.1088/1741-2560/12/3/031001

van Dinteren, R., Arns, M., Jongsma, M. L., & Kessels, R. P. (2014). P300 development across the lifespan: A Systematic review and meta-analysis. *PLoS One, 9*(2), e87347. doi:10.1371/journal.pone.0087347

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

Walter, C., Rosenstiel, W., Bogdan, M., Gerjets, P., & Spüler, M. (2017). Online EEG-based workload adaptation of an arithmetic learning environment. *Frontiers in Human Neuroscience, 11,* 286. doi:10.3389/fnhum.2017.00286

Wang, C. H., Moreau, D., & Kao, S. C. (2019). From the lab to the field: Potential applications of dry EEG systems to understand the brain–behavior relationship in sports. *Frontiers in Neuroscience, 13*, 893. doi:10.3389/fnins.2019.00893

Wang, X.-W., Nie, D., & Lu, B.-L. (2014). Emotional state classification from EEG data using machine learning approach. *Neurocomputing, 129*, 94–106. doi:10.1016/j.neucom.2013.06.046

Wu, Y., Lan, Y., Huang, S., & Lin, Y. (2019). Enhancing medical students’ communicative skills in a 3D virtual world. *Educational Technology & Society, 22*(4), 18-32.

Xu, J., & Zhong, B. (2018). Review on portable EEG technology in educational research. *Computers in Human Behavior, 81*, 340–349. doi:10.1016/j.chb.2017.12.037

Kuromiya, H., Majumdar, R., & Ogata, H. (2020). Fostering Evidence-Based Education with Learning Analytics: Capturing Teaching-Learning Cases from Log Data. *Educational Technology & Society*, *23* (4), 14–29.

Fostering Evidence-Based Education with Learning Analytics: Capturing Teaching-Learning Cases from Log Data

Hiroyuki Kuromiya[1*], Rwitajit Majumdar[2] and Hiroaki Ogata[2]
[1]Graduate School of Informatics, Kyoto University, Japan // [2]Academic Center for Computing and Media Studies, Kyoto University, Japan // khiroyuki1993@gmail.com // dr.rwito@gmail.com // hiroaki.ogata@gmail.com
[*]Corresponding author

(Submitted March 26, 2020; Revised April 23, 2020; Accepted August 27, 2020)

ABSTRACT: Evidence-based education has become more relevant in the current technology-enhanced teaching-learning era. This paper introduces how Educational BIG data has the potential to generate such evidence. As evidence-based education traditionally hooks on the meta-analysis of the literature, so there are existing platforms that support manual input of evidence as structured information. However, such platforms often focus on researchers as end-users and its design is not aligned to the practitioners' workflow. In our work, we propose a technology-mediated process of capturing teaching-learning cases (TLCs) using a learning analytics framework. Each case is primarily a single data point regarding the result of an intervention and multiple such cases would generate an evidence of intervention effectiveness. To capture TLCs in our current context, our system automatically conducts statistical modelling of learning logs captured from Learning Management Systems (LMS) and an e-book reader. Indicators from those learning logs are evaluated by the Linear Mixed Effects model to compute whether an intervention had a positive learning effect. We present two case studies to illustrate our approach of extracting case effectiveness from two different learning contexts – one at a junior-high math class where email messages were sent as intervention and another in a blended learning context in a higher education physics class where an active learning strategy was implemented. Our novelty lies in the proposed automated approach of data aggregation, analysis, and case storing using a Learning Analytics framework for supporting evidence-based practice more accessible for practitioners.

Keywords: Learning analytics, Evidence-based education, Technology-enhanced Evidence-based Education & Learning (TEEL), Learning Evidence Analytics Framework (LEAF), Mixed effects model, Teaching-learning case

1. Introduction

Evidence of good practices in education has been getting prominent around the world (OECD, 2007; European Commission/EACEA/Eurydice, 2017). Governments have developed many applications to store and retrieve evidence in education. For example, What Works Clearinghouse (WWC) by U.S. Department of Education or Evidence Library by UK EEPI Centre were developed to review existing educational researches focusing on the results from a high-quality study. However, since they are often just a database of evidence, users are not supposed to register evidence from their own experience. What users can do is only to search for evidence which is already registered by researchers or educational offices. In some cases, users can register evidence from a form, but users have to input evidence manually, which takes time for users. Currently, practitioners tend to be kept away from the evidence generation process.

To solve these problems, we focus on Learning Analytics approaches. Learning Analytics (LA) is the field which handles educational big data for improving teaching and learning activities. LA has the potential to investigate the impacts of different learning strategies or systems by using students' behavioural logs (Hwang, Spikol, & Li, 2018). In this paper, we propose a LA platform for supporting the evidence generation process for practitioners. Here are the main contributions of the article.

- We propose an automatic case extraction process on a learning analytics platform. Each case is just a description of an intervention, but gathering these cases, they will be evidence of the intervention.
- We show that mixed effects model is useful for case extraction and we develop an interface in our learning dashboard to visualize the results to the users.

In the field of LA, researchers often investigate the evidence of the effectiveness of LA. However, the agenda of how technology can support the process of evidence extraction from big data and building decision support technology is not focused. Hence, we engaged in supporting evidence-based education by using educational big data collected from different learning systems.

2. Related works

2.1. Learning analytics platform to support evidence-based education

To the best of our knowledge, few studies investigate evidence-based practice from a practitioners' perspective. One example is the Learning Analytics Community Exchange (LACE) project's Evidence Hub (Ferguson & Clow, 2016). LACE Evidence Hub followed the evidence-based medicine paradigm to synthesize published LA literature and meta-analyze four propositions about learning analytics - improving learning outcomes, improving learning support and teaching, deployment at scale, and ethics. Evidence Hub collected evidence that will enable the learning analytics community to assess the effectiveness and relative desirability of outcomes resulting from the use of learning analytics tools and techniques. So, the purpose is to collect evidence of the effectiveness of learning analytics practices.

Analytics4Action Evaluation Framework (Rienties et al., 2016) developed at Open University UK is another example. It provides an evaluation flow of the intervention in the context of distance learning education. They map the six key steps in the evidence-based intervention process; (1) determine key metrics while working together with the key stakeholders, (2) decide on a type of intervention, (3) plan an experimental design, (4) analyze outcomes, (5) store evidence into evidence hub, (6) compare different interventions. Through the cycles of these six steps, one illustrates which types of interventions under which conditions have a positive effect and store the results in to OU Evidence Hub. OU Evidence Hub is a branch of LACE Evidence Hub project. In practice, Rienties, Cross, and Zdrahal (2017) investigated the effectiveness of emails and predictive modelling to learners at-risk through the system. In that paper, they examined how evidence-based research design can be implemented to actual learning analytics intervention.

Table 1 summarizes the features of current learning analytics approaches for evidence-based education. Compared to these two studies, our approach is unique in the following points.

- We take evidence as the accumulation of cases. Each case is not itself evidence, but we can gather cases to evaluate the strength of an evidence.
- Our system includes an automated statistical inference step in case extraction while other studies do not automate statistical inference.

We believe these two points are essential for democratizing evidence-based education with the practitioner by harnessing learning analytics tools and techniques.

Table 1. Feature of current learning analytics approach for evidence-based education

	LACE Evidence Hub (Ferguson & Clow, 2016)	Analytics4Action (Rienties et al., 2016)	Proposed system
Evidence from what?	Literature	Designed Experiment	Case Study (Teaching-Learning Case)
How to support Evidence Generation?	Aggregate existing papers	Suggest the evaluation flow of the intervention	Apply statistical models to the data
Statistical calculation	-	Manual	Automated

2.2. Case extraction models

In our system, we aim to estimate the effectiveness of an intervention by automatic adoption of statistical models. At this stage of development, we consider a variety of models that is able to handle the effectiveness of an intervention. We prepared three candidates for case extraction - classical testing method, time-series model, and mixed effects model. Here, we review each statistical model and examine the suitability to our system.

The first candidate is a group comparison approach. As a representative of it, we consider *t*-test as a popular pre-post comparison method. Since it is widely used in the learning analytics field, it is easy to interpret for users. However, there are two disadvantages for our system. First is the information loss by aggregation of the data. Since it does not deal with repeated measurement, we should gather the measurement points by person. The second is that *t*-test has a problem with multiple comparisons where the power of the test decreases if we adopt it many times. As our system allows users to repeat the analysis, it will be a problem in our context.

The second candidate is a time-series model. Here, we consider Interrupted Time Series (ITS) model as a representative of it. Mathematically ITS is the segmented regression model with dummy variables representing the period of the intervention (Bernal, Cummins, & Gasparrini, 2017). In an educational context, Hansen et al. (2014) used ITS model to estimate the impact of a central educational program in German. ITS is a good estimation of the data because learning logs has a time-series structure in nature. However, ITS model compares an intervention period with a baseline, and not necessarily a control period. In ITS, it is difficult to compare two different periods - intervention and control.

The last candidate is a mixed effects model. Mixed effects model is a statistical model that includes a combination of fixed effects and random effects, which represents overall tendency (fixed effects) and individual difference (random effects) for each. It is also used for effectiveness estimation in educational context (Dawson, Jovanovic, Gašević, & Pardo, 2017). For automated extraction of cases from a learning log, such an approach is rational as standard learning logs in any e-learning environment often contain repeated measurement of students. Since it is not a time-series model, it cannot consider the time-series of data. However, instead, mixed effects model can consider students' individual differences. More importantly, it is able to handle two different time periods - intervention and control period and can compare the measurements over those periods.

So far, we discussed the pros and cons of each model. We summarize our discussion in Table 2. According to the table, we concluded that mixed effects model is most appropriate for our purpose. Hence, we adopt a mixed effects model in our current proposed system.

Table 2. Feature of statistical models for case extraction

	Group Comparison Models	Time-series models	Mixed effects model
Example	*t*-test	Interrupted Time-Series	Mixed Effects Model
Advantage	Easy to interpret	Able to capture the time series nature of data	Able to consider students' individual difference
Disadvantage	Information loss by aggregation	Difficult to compare two periods	Cannot handle the time-series behavior in data

3. Our solution: Case extraction system on learning analytics platform

3.1. LA system overview

We had proposed an evidence extraction system integrated into the learning analytics platform as a solution to extract and store evidence from practice (Majumdar, Akçapınar, Akçapınar, Flanagan, & Ogata, 2019; see Figure 1). Teachers can use LEAF in the context of face-to-face learning in higher and secondary education. It has a learning management system *moodle* (see https://moodle.org/), e-book reader *BookRoll* (Ogata et al., 2015), and the dashboard called LAViEW. BookRoll is a learning tracker as well as e-book reader facilitating many learning analytics studies related to the seamless learning environment (Mouri, Uosaki & Ogata, 2018). In our context, teachers conduct an intervention and control method, using these learning tools in their class. We collect data from the Learning Record Store (LRS) for each period – intervention and control. Students' learning log stored in LRS is represented as xAPI format. The accumulated learning log will be an indicator to measure the effect of the intervention in our system.

In LEAF, we have a module which captures teaching-learning activities as a case to generate evidence from it. We call it Teaching-Learning Case (TLC), which is a description of an intervention, which is captured from a single teaching-learning scenario. In LEAF, TLC has a specific format - Context, Problem, Indicator, Intervention, Control, and Results) structure as described in Table 3. This format is modified based on PICO structure, which is commonly used in medicine for making research questions and literature review (Jacobs, 2008). Population, Intervention, Control, and Outcome of Interest corresponds to Context, Solution, Baseline, and Indicator respectively. In addition to PICO format, we added "Problem" and "Results" fields in order to arrange and gather records to make evidence. Although there are other formats of evidence - SPICE framework (Booth, 2006) and PECO format (Morgan, Whaley, Thayer, & Schünemann, 2018), our structure of TLC shares many properties with other evidence formats.

TLC itself is not strong evidence of the intervention because it does not have any strict experimental designs like Randomized-Control Study (RCT). However, by gathering many Teaching-Learning Cases (TLCs), we can generate evidence in the future.

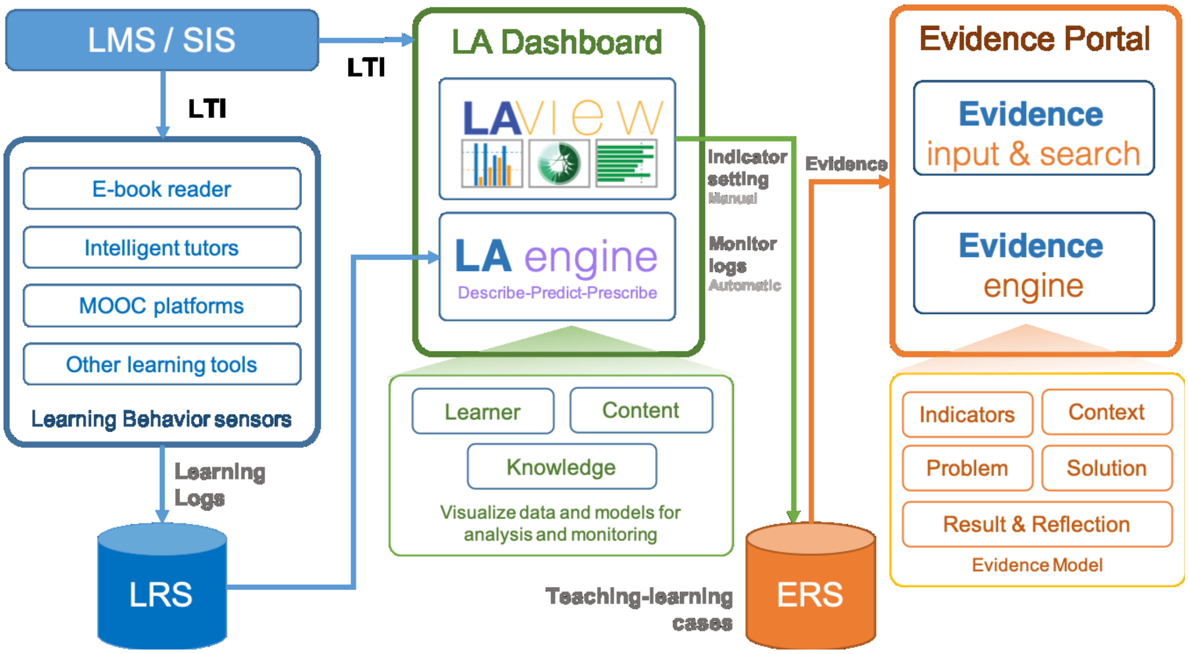


Figure 1. Evidence extraction system on learning analytics platform (Majumdar et al., 2019)

Table 3. The Structure of Teaching-Learning Case in LEAF

Factor	Description	Example
Context course name: subject: grade: class size:	Information regarding the context of evidence. All the information is automatically retrieved from LMS.	course name: Linear Algebra I, subject: Math, grade: B1, class size: 120
Problem	The problems to be addressed in the classroom	Low engagement to homework materials
Indicator	Measurable indicators that a user wants to look at.	Reading time on the materials
Intervention title: dates	The details of the intervention conducted by a teacher. Users need to specify the date of the intervention	title: Send email to students dates: “2019-05-01,” “2019-05-08,” ...
Control title: dates	The details of the control method that a teacher wants to compare with the intervention. The format is the same as the intervention.	title: In-class reminder dates: “2019-04-05,” “2019-04-12,” ...
Results	Description of analysis results by a user. In proposed system, the analysis results are given by the system.	Email intervention increased the reading time by 5.5 min (p = .01)

In this paper, we propose a case extraction module on LEAF. We aim to capture Teaching-Learning Cases from teaching-learning logs stored in Learning Record Store. The characteristics of our system are three listed below.

- Users can estimate the effectiveness of their intervention by just clicking the tool integrated in the learning analytics platform.
- The system provides a database integrated to learning analytics dashboards where users can store and share the effectiveness of their intervention with actual learning context.
- Users can use our system in either online learning environments or face-to-face classroom teaching.

Next, we describe a workflow of capturing TLC in actual teaching-learning context. The statistical aspect of capturing TLC is explained in the last section of this part.

3.2. Workflow of capturing TLC with LEAF

Consider a teacher who teaches their classroom with our learning tools on LEAF. When the teacher conducts a specific intervention, and wants to know the effectiveness of it, our TLC extraction module allows users to register evidence by three steps using the power of educational big data stored in LRS.

The first step is to fill the analysis setting panel (Figure 2A). In this step, users specify the information needed to analyze learning data to conduct the statistical analysis. The course information is automatically retrieved by LTI information, so users do not have to choose. The "Intervention" and "Control" fields are composed by two input fields - name and dates. Users are supposed to write the content of their intervention in name and specify when they conducted the intervention by date picker dialog. Intervention period can be either before or after the control period because we do not consider the order of intervention and control. For source and indicator fields, users can choose the best metric available that fits to measure their problem. Currently, two sources are available - moodle and BookRoll. Considering Moodle as a data source, there are five indicators that can be choose - number of access to moodle contents, moodle pages, moodle quizzes, moodle forum, and external resources. Considering BookRoll as the source, five indicators are available - reading time, number of markers, number of memos, number of bookmarks and number of access to the contents. Once users fill the five fields, push the "Analyze" button to conduct statistical analysis based on the learning log stored in LRS.

The second step is to interpret the results from the system (see Figure 2B and Figure 2C). The overview panel (Figure 2B) represents the inferential statistics estimated by mixed effects model. Overview panel consists of two pieces of information - the effectiveness of the intervention and the strength (reliability) of the case. The effectiveness is retrieved from the coefficients of mixed effects model and the strength of the evidence is determined based on the *p*-value of the coefficient. We refer to the previous studies about the thresholds for p-value interpretation (Raiola & Di Tore, 2012) to categorize p-values into sentences. The strength of the evidence is determined by the following criteria (Table 4). In the example in Figure 2, users can see that the intervention increased the reading time on e-book readers by 11.42 minutes and the results are highly reliable because the strength of the case is "Very Strong." The data panel (Figure 2C) represents descriptive statistics. It compares average reading time over intervention dates and control dates.

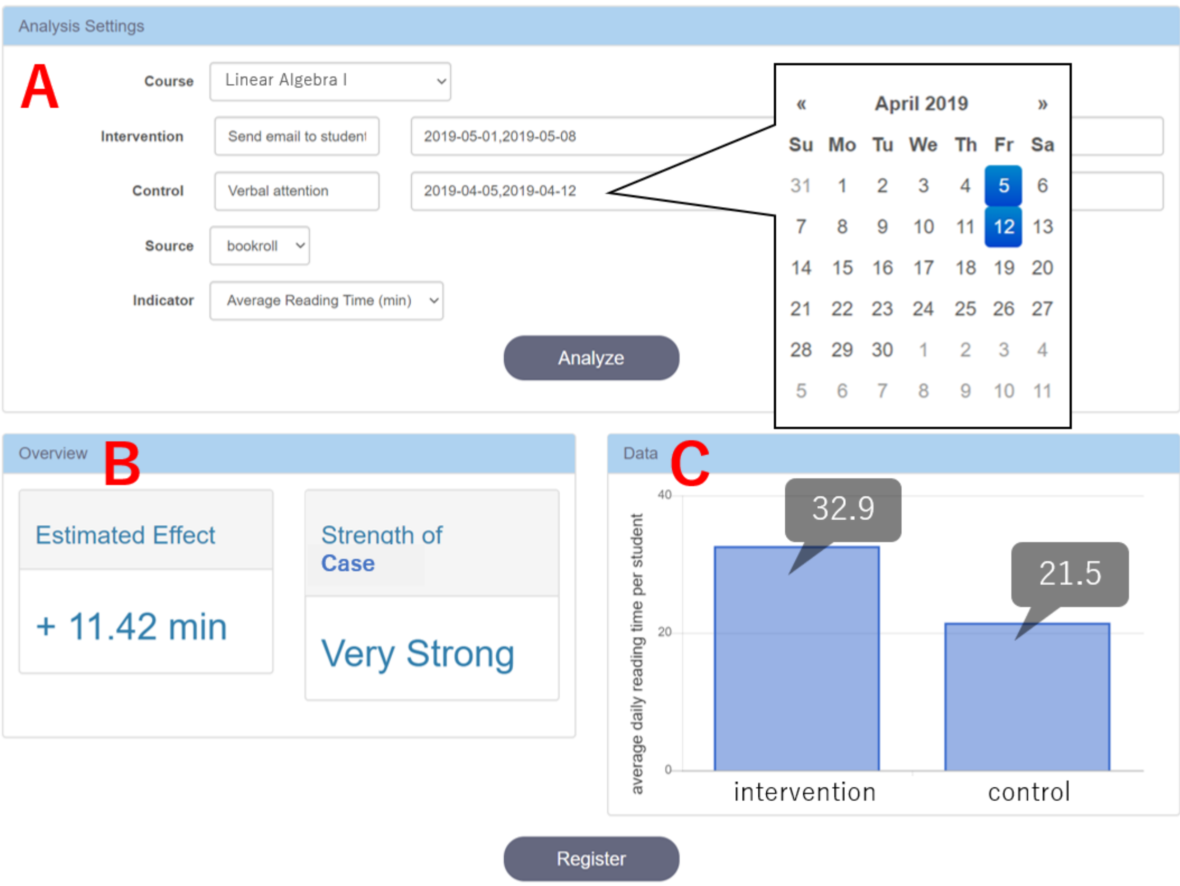


Figure 2. Case extraction page in evidence portal

Once the user interprets the results, users are encouraged to save the results of their intervention in the Evidence Record Store (Figure 3). It has seven input fields – Institution, Course name, Problem, Intervention, Control, Indicator, and Results. It seems to be tough to fill all the fields by manual. However, the fields users have to fill is just one - Problem. Other fields are automatically filled by the system based on the results in the Case Extraction Page. In this example, the population would save the course name “Example Course” from Moodle, Intervention and Control are filled as “Send email to students” and “Teacher-Centered Approach” for each. Indicator is filled as “average reading time on BookRoll.” Results section are written as “estimated effect: 11.42 (p = .00)” because the results from the statistical model are retrieved from the system as well. Users can store additional information such as “subject” or “intervention details” by clicking “Details” to expand the form.

Table 4. Conventional Thresholds for p-value interpretation (Raiola & Di Tore, 2012)

p-value	Interpretation	Strength of case
$p > .1$	Absence of evidence against the null hypothesis	Null
$.05 \le p < .1$	Low evidence against the null hypothesis	Low
$.01 \le p < .05$	Moderate evidence against the null hypothesis	Moderate
$.001 \le p < .01$	Strong evidence against the null hypothesis	Strong
$p < .001$	Very strong evidence against the null hypothesis	Very Strong

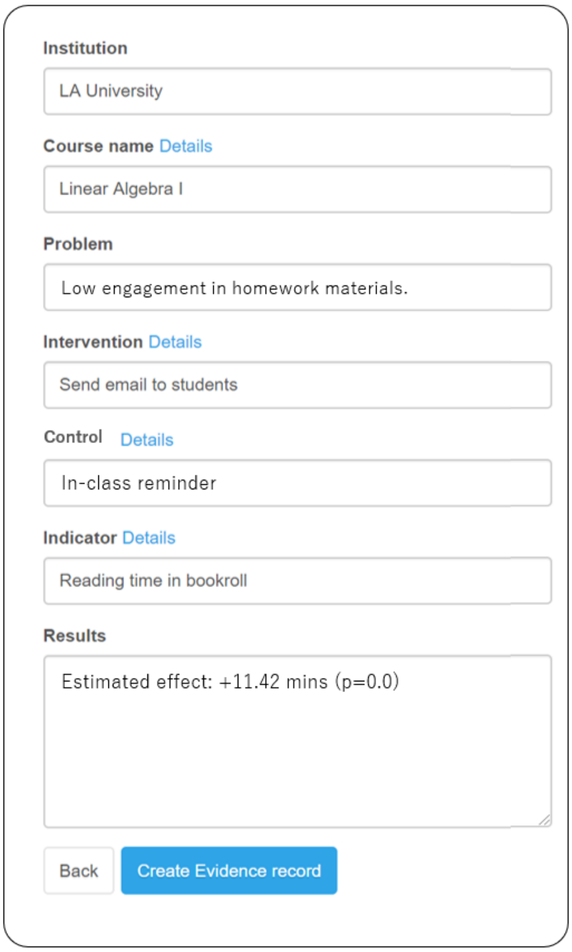


Figure 3. Evidence registration form

Users can see all evidence registered in the Teaching-Learning Cases Table (see Figure 4). It lists every Teaching-Learning Cases with course name, intervention, control, indicator, and results. Users can filter or sort the TLC by subject, grade, or class size. The user will find the cases from similar contexts that match problems that the user has. Users can edit their registered TLC if they get additional results regarding the intervention.

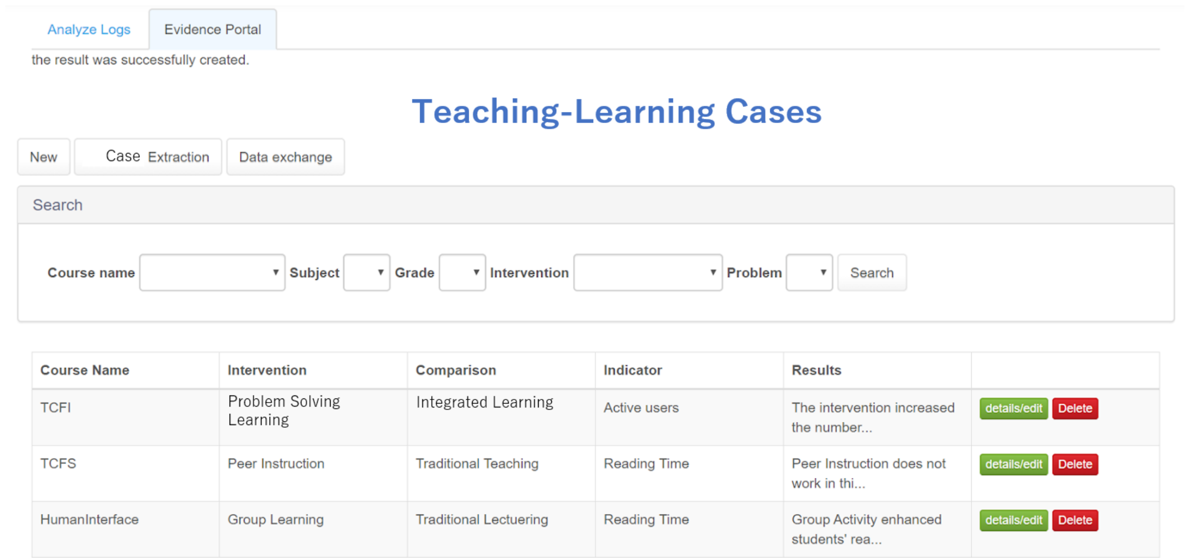


Figure 4. TLC record table

3.4. Evidence extraction process

So far, we have described users' perspective of case extraction. Here we introduce the technical part of our evidence extraction system. It has a four-step process as described in Figure 5. Since we have already introduced result visualization, and case storing step in the previous section, here we describe the details of the first two steps - data collection and modeling and testing step.

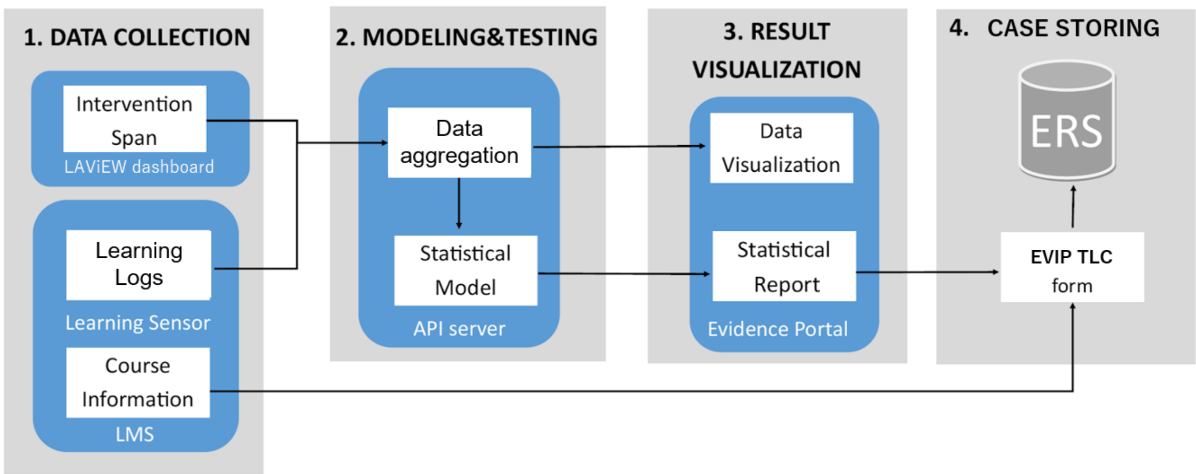


Figure 5. Case extraction process on LEAF

Table 5. Aggregated data structure in data collection step

User_id (*i*)	Date (*j*)	Intervention (*X*)	Indicator (*y*)
221	2019-02-01	0	5.5
221	2019-02-02	0	6.4
...	...	...	...
221	2019-03-01	1	8.4
221	2019-03-02	1	6.9
225	2019-02-01	0	4.1

The first step is the Data *Collection* step where learning records from LRS related to LMS activities or other learning behavior sensor logs are pre-processed in JSON format. The system aggregates the students' log data by each student and day. For example, the following data frame is generated by the system (Table 5). Here, "user_id" represents user identification, "date" denotes for the day of the measurement, "intervention" represents whether it is in the intervention period (1) or control period (0), and "y" denotes the value of outcome of the interest.

The second step is *Modeling and Testing* step, where the system computes statistical inferences about the effectiveness of the intervention based on the collected data. We implemented these computation functions as an external API for evidence extraction. This separation of the application server with evidence extraction server allows expandability of statistical computing. In our current implementation, we use a Python statistical modeling package "statsmodels" (Seabold & Josef, 2010). The specific model parameters used for evidence extraction is described as follows. Currently we have applied a Linear Mixed Effects Model. It fits the hierarchical structure data as represented like above. The modeling approach is described as follows:

$$y_{i,j} = \beta_0 + \beta_1 X_{i,j} + \gamma_{0,i} + \gamma_{1,i} X_{i,j} + \varepsilon_{i,j}$$

where $y_{i,j}$ stands for value of the indicator (the outcome of interest) of student i at date j. $X_{i,j}$ is the dummy variable for the intervention. If the intervention was conducted at time j, $X_{i,j}$ is 1 otherwise 0. β_0 and β_1 represent fixed coefficients ("global parameter") of the model, while $\gamma_{0,i}$ and $\gamma_{1,i}$ do random effects ("local parameter") of the model. At last, $\varepsilon_{i,j}$ is the error term which follows normal distribution. In this case, the fixed effects represent the overall tendency while the random effects show the characteristics of each student in the sample. Since we want to know the overall reaction of students against the intervention, we particularly pay attention to β_1. If β_1 is positive, it means the intervention has a positive effect on the indicator. If it is negative, the intervention has a negative effect on the indicator. At the statistical *Modeling and Testing* step, these coefficients and their p-value were estimated by the EM algorithms (Lindstrom & Bates, 1988).

4. Research study: Use case analysis

Based on our proposed solution in the previous section, we studied the usefulness of the system when actually implemented in a live context. We choose two case studies with two different types of data sources on LEAF. In one case (Case A) it has the learning log from an e-book reader, and in the other case (Case B) it has the log data from LMS. The two cases also focus on classrooms in two different countries and for different educational contexts. Thus, the two cases provide context variation to demonstrate to our current reader the LEAF system's application. As we consider the log data, we highlight the use case to extract effectiveness in terms of increased engagement as measured by the interaction logs.

In each case, the teacher conducted an intervention, and log-data was collected during the experiment through LEAF system. After the intervention, we used the system, which applied the linear mixed effects model and saved our interpretation as the results. Of the two cases presented, Case A checks the effectiveness of an e-mail intervention in junior high school where students were reminded about preview and review activities in an e-book reader. The data source here was the students' reading logs retrieved from the e-book reader. Case B evaluates the effectiveness of an active learning strategy, Peer Instruction, in an undergraduate physics class. The data source was Moodle logs. Our system could be used to capture the effectiveness of Peer Instruction in comparison to traditional teaching in that context.

4.1. Case A: Extracting effectiveness of E-mail reminders to boost engagement of Junior-High School Students

Our first case concerns a math class at a Junior-high school in Kyoto, Japan. We targeted 60 first-grade students and conducted the intervention to enhance students' learning engagement. The teacher sent e-mail messages to the students during the intervention period. The intervention was conducted on 4, 5, 6, and 9 December 2019. The control period (no e-mail messages were sent) was 8, 9, 10, 14 January 2020. The course topics in the experimental period were "data utilization" in statistics and "polynomial" in algebra. The reason we selected these topics was that these topics were taught by traditional lecture format which does not include much of group works, or interactive sessions. The teacher used the e-book reader in the class to upload the learning materials and create quizzes within that material. Students can then annotate with a maker and write memos while reading and attempt the set quizzes within the same environment. In this case, the teacher uploaded the math textbook and other learning materials such as teacher's notes, practice problems and its solutions. The objective of our analysis was to explore how effective e-mail messages were to enhance students' learning activities.

4.1.1. Analysis settings

First, we filled the class activity schedules. During the intervention period, the teacher sent six messages to students. The teacher noted words of encouragement, as well as the essential information, was part of the email message. Five out of six messages contained the notification of the next class's requirement, with an assumption that it would encourage students to preview and review the learning materials outside of the class. Each email was sent to all students. As the purpose of the intervention was to encourage students' previewing and reviewing motivation, the indicator we used was the average reading time. The logs which exceed 20 minutes were automatically removed by the system. It removed the data where students were just keeping an eBook open. Analysis settings for this case are shown in Figure 6A. To validate our results from the system, we conducted a brief survey to the students in order to investigate students' perception of the intervention. We asked students that (1) did you feel easy to receive email messages from the teacher? (response options: very easy, easy, not so easy, not easy) and (2) Did you think it was useful for your learning to receive the email message from the teacher? (response options: very useful, useful, not so useful, not useful).

4.1.2. Case extraction results

Figure 6B and Figure 6C represents the analysis results on the case extraction page. The overview panel (Figure 6B) showed the effectiveness of the intervention. We have attached a table of the analysis results as well in order to check the accuracy of it (Table 6). This table is not shown on the screen but logged in our analysis API server whenever the user conducts an analysis. The panel shows that email intervention increased the reading time on e-book readers by 4.74 minutes compared to the control period with moderate strength of case. Checking against the table, it does match to the results of actual mixed effects model results. The data panel (Figure 6C) shows the descriptive mean reading time both in intervention and control period. It indicates that the reading time on the e-book reader was longer (average: 27.42 min) in the intervention period than that in the control period (average: 22.77 min). For the survey response, 56 out of 60 students (93%) responded to the survey. Among the responders, 52 students (93%) responded that receiving email messages were easy (very easy or easy) and 27 students (48%) thought that the messages were useful (very useful or useful) for their learning at home.

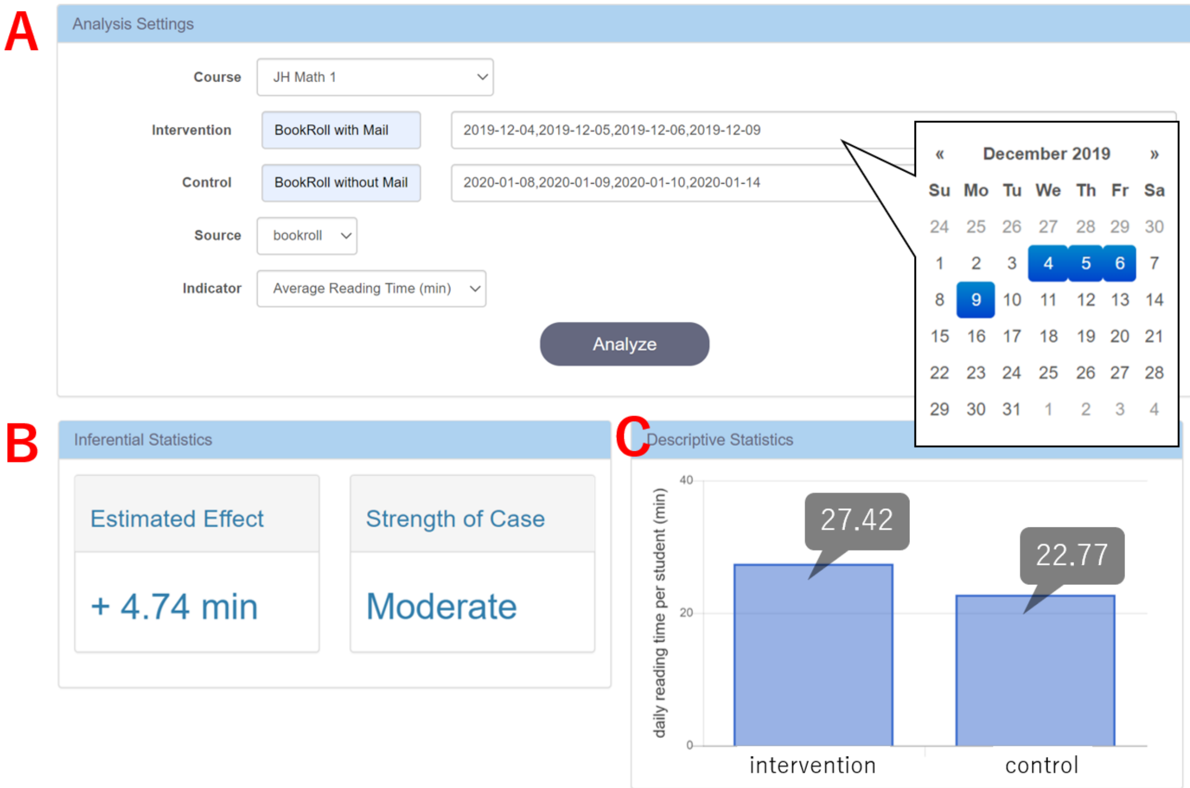


Figure 6. Analysis results on study case: Effectiveness of e-mail messages

Table 6. Mixed linear model regression results

	Coefficient	Standard Error	*p*-value
Intercept	22.55	1.73	.00
Intervention	4.74	2.17	.029
Group Variance	10.11	0.76	-

4.1.3. Findings from Case A

The results suggested that our system can be adopted to a traditional teaching-learning class in secondary education. As our platform offers a learning management system and e-book reader with learning analytics dashboard, teachers can quickly adopt technology-enhanced learning to their class. It produces a potential to extract the effectiveness of an intervention from teaching-learning logs in a face to face classroom situation.

It is beneficial for teachers because they can explore the effectiveness of the intervention through our system. In this case, although the effectiveness of email intervention to students' learning engagement were observed in many studies (Arnold & Pistilli, 2012; Heiman, 2008), few studies deal with the situation where instructors send a message via the learning analytics system to students at the junior high school level in Japan. It is often the case that a method that worked well elsewhere would be completely ineffective elsewhere.

Apart from the quantified change measured by the system, we collected the perceived improvement from the point of view of the teacher and also that of the students. Both the teacher and portion of the students perceived the message boasted their learning engagement. While such perception confirms the face validity of the system's result as verified by the stakeholders, we plan for formal validity and reliability evaluation in our future studies.

4.2. Case B: Extracting effectiveness of peer instruction strategy in a blended learning context

The second case was an introductory Physics course for first-year undergraduate students. The course instructor used LEAF platform to offer the semester-long course from 18 January to 8 April 2019. There were 64 students enrolled in the class, out of which 59 registered on the Moodle. The instructor adopted a specific in-class active learning strategy, Peer Instruction for a particular topic and Traditional lecturing for the next topic. The course was taught in a blended-learning environment using the LEAF platform. The teacher adopted LCM (Learner-Centric MOOCs) model to their class for online activities. LCM model is a prescriptive (pedagogical) model consisting of a set of guidelines, activity formats and actions for MOOC creators (Murthy, Warriem, Sahasrabudhe, & Iyer, 2018). LCM model consists of four learning components (see Figure 7). Here are typical examples of each component.

- Learning Dialogs (LeD): short videos with reflection spots
- Learning by Doing (LbD): multiple choice questions with customized and constructive feedback
- Learning Experience Interaction (LxI): peer learning through focused discussion
- Learning Extension Trajectories (LxT): various materials to address the diversity of learners

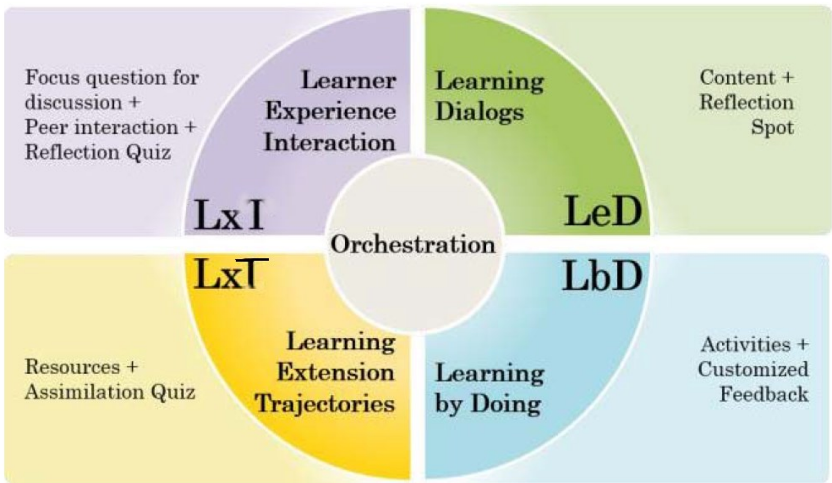


Figure 7. Four components of LCM model (Murthy et al., 2018)

In this paper, we will measure the effectiveness of Peer Instruction based on these components. We will use “the number of access to moodle contents” for LeD, “the number of access to moodle quizzes” for LbD, “the number of moodle forum” for LxI, and “the number of external resources” for LxT.

Previously the instructors reported a reflective practitioner study to elaborate the design decisions taken for the blended course activities (Kannan & Gouripeddi, 2019). We reported our initial approach of analysis and how it can help to evaluate a pedagogical model using learning logs (Kuromiya, Majumdar, Warriem & Ogata, 2019). In this paper, we further investigate the effectiveness of Peer Instruction in LCM context based on four learning components with the help of our case extraction system.

4.2.1. Analysis settings

Analysis Settings were filled like in Figure 8A. The teacher conducted the Peer Instruction method on 24, 25, 26, 27 February, and 6,7,8, 11, 12 March 2019. For comparison, the teacher held a traditional lecturing on 13, 14, 18, 19, 20, 21, 22 March 2019. As the purpose of the intervention was to enhance students’ engagement, we selected four indicators which correspond to four components in the LCM model as we described in the previous section. We selected “number of access to moodle contents,” “number of access to moodle quiz,” “number of access to moodle forum,” and “number of access to moodle resources.” Each indicator corresponds to LeD, LbD, LxI, LxT components for each.

4.2.2. Case extraction results

In this case, we have conducted four analyses as mentioned. Along with it, we showed the results from “number of access to moodle contents (LeD)” in Figure 8. From the overview panel (Figure 8B), we can see that the Peer Instruction increased the number of access to moodle contents by 0.43 times. However, we can also see that the strength of the evidence was very weak, which means the results were not reliable. The result from the data panel (Figure 8C) suggested that the number of access to moodle contents in the intervention class was slightly higher than that in the control period. The average number of access to moodle contents in the intervention period was 3.57, while that in the control class was 3.17. The results from other three indicators were summarized in Table 7.

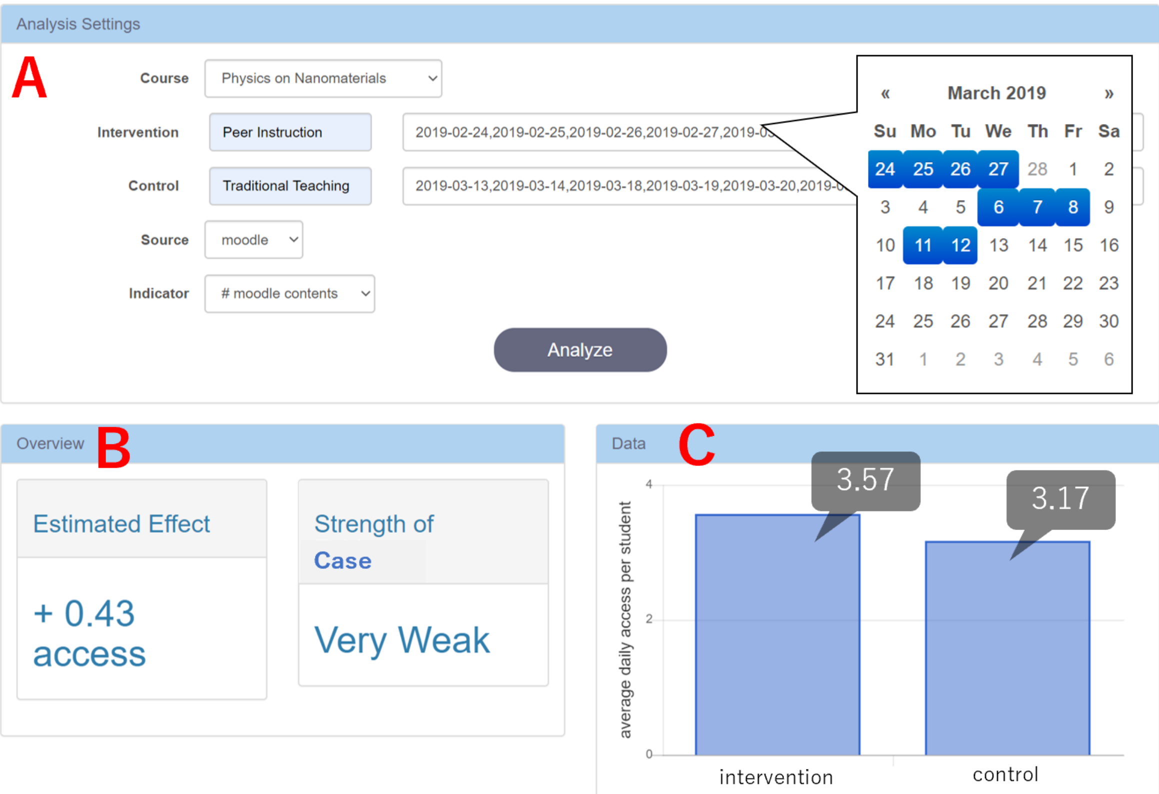


Figure 8. Analysis results on Case B: Effectiveness of the peer instruction in blended learning course

Table 7. Analysis results for each indicator

Indicator	Estimated Effect		Strength of Case
Moodle Contents (LeD)	+	0.43	Null
Moodle Quiz (LbD)	+	5.73	Low
Moodle Forum (LxI)	+	0.35	Null
External Resources (LxT)	-	1.45	Null

The results showed that Peer Instruction had no evidence of the effects to increase the number of activities on moodle compared to Traditional Teaching method. According to Table 7, only to the moodle quiz (LbD) the intervention had relatively positive effects. However, the strength of the evidence was low, which means the results were not so reliable. It suggests that we cannot say anything whether Peer Instruction activity had increased student engagement in blended online learning settings in this case.

4.2.3. Findings from Case B

This case can be considered as a demonstration of how the case extraction system was used to analyze class activity during a blended learning course. In blended learning courses, most students' learning logs are collected from the learning management system. Previous literature (Lu, Huang, Huang, Lin, Ogata, & Yang, 2018) pointed out that students' final academic performance could be predicted from only one-third of the log data in blended learning context. We can capture students' learning activities and infer the effectiveness of the intervention only from accumulated learning log by adopting our system. Also, teachers can investigate the effectiveness of the intervention from multiple aspects by selecting different indicators or sources. It is one of the advantages of using our case extraction system.

Previous literature reports that Peer Instruction encourages students' engagement in face-to-face classroom (Busebaia & John, 2020) and the online learning environment (Liu, 2019). However, no studies consider a blended learning environment. Particularly, the teacher used a specific type of blended learning strategy, Learner-Centric MOOCs (LCM) model to design the learning activities for their class. Such a combination of Peer Instruction and activities based on LCM model was not investigated yet.

5. Discussion and conclusion

5.1. Contributions to current learning analytics systems

Current learning analytics systems do not have a module for intervention evaluation. Although there are some studies where researchers tried to collect evidence about learning analytics interventions (Papamitsiou & Economides, 2014), there are no systematic solutions to collect evidence from log data. In our system, users to do reflections anytime they want by extracting the corresponding data and executing the statistical computation. It enables teachers to analyze their interventions almost in real-time. It is useful in practice because sometimes teachers have to decide whether they should continue an intervention or not. Moreover, they can retry analysis changing parameters - indicators or sources. It gives them opportunities to find unexpected things and to investigate their interventions deeper by looking at it from various aspects.

Moreover, our two case studies showed that our extracted case was essential for future teaching practice in the context where previous literature is insufficient to compare effectiveness. Traditionally, evidence-based education is for researchers and policymakers, based on well-designed RCTs. Herodotou, Heiser, and Rienties (2017) discussed the possibilities of adopting RCTs in learning analytics fields for making better evidence. Often it would be challenging to involve the teachers directly in the evidence-based practice. Our approach suggested a new type of evidence-based education based on data generated during the teaching-learning intervention which the teachers themselves orchestrate. Applying learning analytics techniques, we demonstrated how to automate data collection, analysis, and case storing for supporting evidence-based practice. It democratizes evidence-based education to practitioners and aims to make data-informed teaching practice more accessible.

5.2. Validity of our data analysis process

In both case studies, we used the indicators which can be easily retrieved from students' learning log. As we showed in case studies, these parameters can serve as a good measure for students' learning engagement.

However, we also understand that many teachers are interested in their performance as well as their learning activities. Teachers often conduct interventions in order to improve students' performance rather than their learning engagement. In that context, existing indicators are not enough. We plan to add indicators regarding students' performance such as quiz scores or exam scores.

For the *Modeling* step, we used a mixed effect model to estimate the effectiveness of the intervention in the case. By using a mixed effect model, we could consider students' differences and offer the interface which is easy to understand for users. However, we need to add models in order to deal with new types of data such as performance scores. Our project is running in the open-source spirit; we would like to allow users to define their own models for case extraction. For instance, Open Strategic Data Projects hosted by the Center for Education Policy Research, Harvard University, makes analytic tools freely available for educational data analysts (Passmore & Chae, 2019). Many analysis models published there serve as good candidates of case extraction models in our system for different analysis.

For the *Testing* step, we should mention the multiple comparison problem. Since our system allows users to do multiple adoptions of a statistical model to different indicators, users may fall into false-positives based on this repetition. That is why we decided not to use p-value as statistical testing. In both cases, we refer p-value as the strength of the case. It does not determine if the intervention worked or not but offers gradual confidence in the results. According to the statement from American Statistics Association, they noted that "Scientific conclusions and business or policy decisions should not be based only on whether a *p*-value passes a specific threshold" (Wasserstein & Lazar, 2016, p. 131). Our approach to p-values meets this principle as we do not care about if the *p*-value exceeds a significance level or not.

5.3. Limitations

Compared to the previous studies - LACE Evidence Hub (Ferguson & Clow, 2016) and Analytics4Action (Rienties et al., 2016) (see Table 1), our system is unique to apply statistical computing automatically with learning logs. However, there are still some limitations regarding reliability and flexibility in our current approach. As our platform is primarily focused on easing the data synthesis and the statistical computation within a context using existing Learning Analytics tools, the actual selection of indicators, the context and the interpretation still remains as inputs by users. Currently the system does not evaluate the validity of that user interpretation and recorded result. However, for instance in LACE Evidence Hub, it registers only reliable meta-analysis of the literature. Similarly, regarding system flexibility, currently our system is able to extract students' engagement only based on students' behavior online. Rienties et al. (2017) pointed out that there are three levels in possible impacts of learning analytics systems - attitudes, behavior, and cognition. While Analytics4Action can collect all three types of data, our system can potentially collect all of them in the framework, however now only behavioral level data is automatically synthesized. Despite these limitations, we believe that there are benefits in democratizing evidence-based education with the practitioner by harnessing learning analytics tools and techniques to extract evidence of learning from log data.

5.4. Future research direction

In this paper, we proposed a semi-automated analysis system to the log data stored on a learning analytics platform. However, it does not guarantee the quality of the case. We only provided a place to do calculations for case registration. We should investigate our system further focusing on what context under what conditions the intervention works. Hence, we plan to develop an additional module to aggregate various teaching-learning cases into evidence. Figure 9 shows the evidence generation process on our case extraction system. By aggregating (meta-analyzing) similar teaching-learning cases, our system will be able to produce more robust evidence of the intervention.

Furthermore, we are planning to implement a case recommendation function in Evidence Portal. Once the case records keep logging in the Evidence Record Store, it might be difficult for users to find the best solution that matches their problems and situation. Case recommendation function offers the most relevant solution according to the users' context and problem. There we shall compute the context similarity between cases from the context field in the case records. It will promote users to refer to past case records rather than only to rely on their intuitive practice.

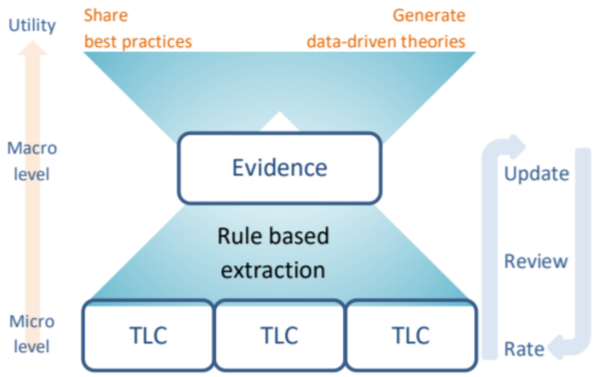


Figure 9. Evidence generation from teaching-learning cases

5.4. Conclusion

In this paper, we proposed an integrated case-extraction system based on a learning analytics platform. We showed that (1) the data processing flow of case extraction from students' learning log, (2) Mixed effects model can be used for automated case extraction, and (3) Our extraction module implemented on LEAF makes evidence-based practice accessible for practitioners by showing actual case studies.

While many implemented LA dashboards only visualize the descriptive statistics, our system design pushes that boundary to show the result of statistical modelling to users on the learning analytics platform itself. We demonstrated with two case studies of our system during a live teaching-learning scenario, where a statistical model was used to extract the effectiveness of pedagogical interventions conducted in those scenarios. In the future, we plan to build an additional module to extract evidence by meta-analyzing the cases extracted by the proposed system. We believe that integration of learning analytics techniques to automate log-data analysis, with the stakeholder's workflow has a potential to democratize evidence-based practice for different stakeholders beyond only researchers.

Acknowledgement

This research was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (S) Grant Number 16H06304, NEDO Special Innovation Program on AI and Big Data 18102059-0, JSPS KAKENHI Grant-in-Aid for Research Activity Start-up Grant Number 19K20942, JSPS KAKENHI Grant-in-Aid for Early-Career Scientists 20K20131 and SPIRITS 2020 of Kyoto University. We sincerely thank all the students from the school in Japan and university in India. Especially, we thank Mr. Go Miyabe, Ms. Saki Maruyama, Ms. Mizuki Yamazaki, and Dr. Vijaya Nandhini Kannan, the instructors of the courses who kindly participated in our study.

References

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267–270). doi:10.1145/2330601.2330666

Bernal, J. L., Cummins, S., & Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: A Tutorial. *International Journal of Epidemiology*, *46*(1), 348–355.

Booth, A. (2006). Clear and present questions: formulating questions for evidence based practice. *Library Hi Tech*, *24*(3), 355–368.

Busebaia, T. J. A., & John, B. (2020). Can flipped classroom enhance class engagement and academic performance among undergraduate pediatric nursing students? A Mixed-methods study. *Research and Practice in Technology Enhanced Learning*, *15*(1), 1-16. doi:10.1186/s41039-020-0124-1

Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2014). Mechanisms and the evidence hierarchy. *Topoi*, *33*(2), 339–360. doi:10.1007/s11245-013-9220-9

Dawson, S., Jovanovic, J., Gašević, D., & Pardo, A. (2017). From prediction to impact: Evaluation of a learning analytics retention program. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 474–478). doi:10.1145/3027385.3027405

European Commission/EACEA/Eurydice. (2017). *Support Mechanisms for Evidence-based Policy-Making in Education. Eurydice Report*. Luxembourg: Publications Office of the European Union.

Ferguson, R., & Clow, D. (2016). Learning analytics community exchange: Evidence hub. In *Proceedings of the 6th International Learning Analytics and Knowledge (LAK) Conference* (pp. 520–521). doi:10.1145/2883851.2883878

Hansen, S., Schwab, F., Schneider, S., Sohr, D., Gastmeier, P., & Geffers, C. (2014). Time-series analysis to observe the impact of a centrally organized educational intervention on the prevention of central-line-associated bloodstream infections in 32 German intensive care units. *The Journal of Hospital Infection*, *87*(4), 220–226.

Heiman, T. (2008). The Effects of e-mail messages in a distance learning university on perceived academic and social support, academic satisfaction, and coping. *Quarterly Review of Distance Education*, *9*(3), 237-248.

Herodotou, C., Heiser, S., & Rienties, B. (2017). Implementing randomised control trials in open and distance learning: A Feasibility study. *Open Learning: The Journal of Open, Distance and E-Learning*, *32*(2), 147–162.

Hwang, G.-J., Spikol, D., & Li, K.-C. (2018). Guest editorial: trends and research issues of learning analytics and educational big data. *Educational Technology & Society*, *21*(2), 134–136.

Jacobs, S. K. (2008). Research guides: Health (Nursing, Medicine, Allied Health): Literature reviews. *Oncology*, *41*, 1302.

Kannan, V., & Gouripeddi, S. P. (2019). Contextualising the learner-centric MOOCs model for effective blending of flipped-classroom method in engineering physics course. In *Proceedings of the 2019 IEEE tenth international conference on technology for education (T4E)* (pp. 46–53). doi:10.1109/T4E.2019.00-51

Kuromiya, H., Majumdar, R., Warriem, J., & Ogata, H. (2019). Data-driven validation of pedagogical model - A Case of blended LCM model. In *Proceedings of the 2019 IEEE Tenth International Conference on Technology for Education (T4E)* (pp. 38–45). doi:10.1109/T4E.2019.00016

Lindstrom, M. J., & Bates, D. M. (1994). Correction to Lindstrom and Bates (1988): Newton-Raphson and EM algorithms for linear mixed effects models for repeated measures data. *Journal of the American Statistical Association, 89*(428), 1572-1572. doi:10.2307/2291042

Liu, Y. (2019). Using reflections and questioning to engage and challenge online graduate learners in education. *Research and Practice in Technology Enhanced Learning*, *14*(1), 3. doi:10.1186/s41039-019-0098-z

Lu, O. H. T., Huang, A. Y. Q., Huang, J. C. H., Lin, A. J. Q., Ogata, H., & Yang, S. J. H. (2018). Applying learning analytics for the early prediction of students' academic performance in blended learning. *Educational Technology & Society*, *21*(2), 220–232.

Majumdar, R., Akçapınar, A., Akçapınar, G., Flanagan, B., & Ogata, H. (2019, March). *Learning analytics dashboard widgets to author teaching-learning cases for evidence-based education*. Paper presented at the 9th International Learning Analytics and Knowledge (LAK) Conference, Tempe, AZ. Retrieved from http://hdl.handle.net/2433/244167

Morgan, R. L., Whaley, P., Thayer, K. A., & Schünemann, H. J. (2018). Identifying the PECO: A Framework for formulating good questions to explore the association of environmental and other exposures with health outcomes. *Environment International*, *121*(Pt 1), 1027–1031.

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

Murthy, S., Warriem, J., Sahasrabudhe, S., & Iyer, S. (2018). LCM: A Model for planning, designing and conducting learner-centric MOOCs. In *Proceedings of the 2018 IEEE Tenth International Conference on Technology for Education (T4E)* (pp. 73-76). doi:10.1109/T4E.2018.00022

Organisation for Economic Co-operation and Development (OECD). (2007). *Evidence in education: Linking research and policy*. Paris, France: OECD Publishing.

Ogata, H., Majumdar, R., Akçapinar, G., Hasnine, M. N., & Flanagan, B. (2018). Beyond learning analytics: Framework for technology-enhanced evidence-based education and learning. In *Proceedings of the International Conference on Computer in Education (ICCE 2018)* (pp. 493–496). Manila, Philippines: Asia-Pacific Society for Computers in Education (APSCE).

Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). E-Book-based learning analytics in university education. In *Proceedings of the International Conference on Computer in Education (ICCE 2015)* (pp. 401–406). China: Asia-Pacific Society for Computers in Education.

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

Passmore, D. L., & Chae, C. (2019). Potential for Meta-Scientific inquiry to improve the usefulness of HRD research outcomes for practice. *Advances in Developing Human Resources*, *21*(4), 409–420. doi:10.1177/1523422319870173

Raiola, G., & Di Tore, P. A. (2012). Statistical study on bodily communication skills in volleyball to improve teaching methods. *Journal of Human Sport and Exercise*, *7*(2), 468–488. doi:10.4100/jhse.2012.72.12

Rienties, B., Boroowa, A., Cross, S., Kubiak, C., Mayles, K., & Murphy, S. (2016). Analytics4Action evaluation framework: A Review of Evidence-Based learning analytics interventions at the open university UK. *Journal of Interactive Media in Education*, 2016(1). doi:10.5334/jime.394

Rienties, B., Cross, S., & Zdrahal, Z. (2017). Implementing a learning analytics intervention and evaluation framework: What works? In B. Kei Daniel (Ed.), *Big data and learning analytics in higher education: Current theory and practice* (pp. 147–166). Cham, Switzerland: Springer International Publishing. doi:10.1007/978-3-319-06520-5_10

Seabold, S., & Josef P. (2010). Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference* (Vol. 57, pp. 92-96). Retrieved from https://pdfs.semanticscholar.org/3a27/6417e5350e29cb6bf04ea5a4785601d5a215.pdf

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician*, *70*(2), 129–133.

Lin, Y.-T., Wu, C.-C., Chen, Z.-H., & Ku, P.-Y. (2020). How Gender Pairings Affect Collaborative Problem Solving in Social-Learning Context: The Effects on Performance, Behaviors, and Attitudes. *Educational Technology & Society*, *23* (4), 30–44.

How Gender Pairings Affect Collaborative Problem Solving in Social-Learning Context: The Effects on Performance, Behaviors, and Attitudes

Yu-Tzu Lin[1,2], Cheng-Chih Wu[1,2*], Zhi-Hong Chen[1,2] and Pei-Yi Ku[1]
[1]Graduate Institute of Information and Computer Education, National Taiwan Normal University, Taiwan //
[2]Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, Taiwan //
linyt@ntnu.edu.tw // chihwu@ntnu.edu.tw // zhchen@ntnu.edu.tw // smile.peggy333@gmail.com
[*]Corresponding author

(Submitted April 19, 2020; Revised May 1, 2020; Accepted August 26, 2020)

ABSTRACT: This study aimed to investigate the effects of gender pairings on collaborative problem-solving performance, processes, and attitudes in a social learning context. Three types of pairings (i.e., male-male, female-female, and mixed pairings) were considered in an empirical study with 222 tenth-grade students. The selection of three different schools facilitated discussions regarding which schools were more divergent and competitive in a social learning context. The students were asked to solve computer science problems on a social media platform. The results revealed that (1) the single-gender groups had more focused discussions than the mixed-gender groups. Specifically, the male-male groups tended to develop and test their solutions directly without spending significant time on problem identification. Consequently, the single-gender groups exhibited superior performance compared to the mixed-gender groups in terms of applying their knowledge to problem solving. In terms of attitudes toward social learning, the female-female groups were more attentive to the benefits of social learning than the male-male groups. (2) The mixed gender groups had more diverse and divergent discussions compared to the single-gender groups. The educational implications of these findings are also discussed in this paper.

Keywords: Social learning, Gender pairing, Gender difference, Social media, Problem solving

1. Introduction

Based on the development and expansion of social media platforms, their usefulness in the field of education has received increased attention (Ranieri, Manca, & Fini, 2012; Al-Rahmi, Alias, Othman, Marin, & Tur, 2018; Moghavvemi, Sulaiman, Jaafar, & Kasem, 2018; Prestridge, Tondeur, & Ottenbreit-Leftwich, 2019). Because social media provides an interactive social learning environment, by forming connections between heterologous learning communities, students have more opportunities to exchange ideas and opinions with others, which helps foster knowledge construction (Kimmerle, Moskaliuk & Cress, 2011) and has a significant impact on learning performance (Puntambekar, 2006; Moskaliuk, Kimmerle, & Cress, 2012). Significant research efforts have been devoted to investigating how students interact and how their interactions influence learning (Liu & Tsai, 2008; Hou & Wu, 2011; Avcı Yücel & Usluel, 2016; Moghavvemi, Sulaiman, Jaafar, & Kasem, 2018). Such research contributes to developing effective instructional methods with respect to social learning (Moreno, Gonzalez, Castilla, Gonzalez, & Sigut, 2007; Al-Rahmi, Alias, Othman, Marin, & Tur, 2018) and provides opportunities for social learning and interaction, which can help students develop new perspectives by comparing and integrating divergent and conflicting opinions (Moskaliuk, Kimmerle, & Cress, 2012).

However, a social network involves a complex composition of individuals and their relationships, and these factors may influence knowledge sharing and learning performance (Eid & Al-Jabri, 2016). Therefore, team dynamics should be considered in social learning environments (Heo, Lim, & Kim, 2010; Kimmerle, Moskaliuk & Cress, 2011; Lin, Hou, Wang, & Chang, 2013; Avcı Yücel & Usluel, 2016; Son, Kim, Na, & Baik, 2016; Wong & Hsu, 2016). Individual differences are one of the main factors influencing social learning. Student characteristics and interaction styles (e.g., densities, positions, and ties in the social network) influence learning performance in different ways (Chung & Paredes, 2015; Wong & Hsu, 2016). Therefore, various grouping strategies considering individual differences have been considered to facilitate social learning, such as heterogeneous and homogeneous strategies. Many studies have advocated the use of heterogeneous groupings to leverage the diversity of student abilities and characteristics (Wilkinson & Fung, 2002) because heterogeneity can provide additional opportunities to learn from peers with different ability levels and backgrounds. Additionally, various pairing strategies can promote positive learning interactions for collaborative problem solving and improve learning performance (Chen & Chang, 2014).

Regarding pairing strategies, gender differences represent a significant factor that should be concerned in a social learning environment (Underwood, Underwood, & Wood, 2000; Prinsen, Volman, & Terwel, 2007; Zhan, Fong, Mei, & Liang, 2015). Some research has indicated that mixed-gender groups engage less in collaborative problem solving and develop fewer solutions compared to single-gender groups (Underwood, Underwood, & Wood, 2000). However, other research has indicated that mixed-gender pairs provide more divergent knowledge elaboration patterns, which is a significant predictor for learning performance (Ding, Bosker, & Harskamp, 2011). Additionally, males and females may perform differently in different gender pairings. For example, males tend to be more active in dominating the discourse in mixed-gender groups (Prinsen, Volman, & Terwel, 2007) and are more deliberate in terms of problem solving in mix-gender groups compared to single-gender groups. Female-only groups tend to share more personal opinions (Savicki, Kelley, & Oesterreich, 1998), distribute more information for problem solving (Ding & Harskamp, 2006), and are typically more satisfied with social learning (Savicki, Kelly, & Lingenfelter, 1996). Our review of Scopus journal articles published over the past decade indicated that most studies have limited their social learning scenarios to the context of a class or a school. In a larger social learning environment, the effects of group dynamics may be more complex and perceptible. Therefore, larger context requires additional exploration.

This study aimed to clarify how gender parings affect collaborative problem solving in a social learning context by analyzing student learning behaviors during social learning, which can contribute to the educational research field in terms of learning contexts and methodologies. Regarding learning contexts, previous studies have largely focused on discussing grouping strategies for collaborative learning in classroom environments. This study extended the learning context from a classroom to a cross-school social learning platform. Regarding methodologies, most previous studies have applied statistical or qualitative approaches. This study adopted both sequential analyses and statistical approaches to investigate the dynamic and sequential behaviors of students during collaborative problem solving to determine the impact of gender pairings on learning. We attempted to answer three major research questions.

- How do gender pairings influence student performance in terms of collaborative problem solving in a social learning context?
- How do gender pairings influence student behaviors during collaborative problem solving in a social learning context?
- How do gender pairings influence student attitudes toward social learning?

2. Literature review

2.1. Social learning and grouping strategies

In a social learning environment, students have more opportunities to exchange ideas and opinions with others, which helps foster knowledge construction (Kimmerle, Moskaliuk & Cress, 2011). Social learning theory (Bandura, 1977) suggests that individuals construct ideas by observing and imitating peer behaviors. Social constructivism (Vygotsky, 1978) claims that knowledge is constructed through interactions between individuals. The divergence and incongruity observed in social learning environments have been shown to have a significant impact on learning performance (Moskaliuk, Kimmerle, & Cress, 2012; Puntambekar, 2006). Enhancing communication and interaction can help elucidate the problems being discussed and facilitate the process of creativity (Al-Zahrani, 2015). Students who have many active bidirectional interactions with their peers tend to perform better in terms of collaborative problem solving (Chen & Chang, 2014).

Research on the implementation of effective social learning has received significant attention in recent years (Moghavvemi, Sulaiman, Jaafar, & Kasem, 2018). However, appropriate methodologies for utilizing social media in education are still under active investigation. Additionally, although previous research has indicated the benefits of cross-culture social learning, such as cross-organization or cross-country learning (Berkes, 2009; Douglas, Farley, Lo, Proskurowski, & Young, 2010), there has been a lack of research exploring the implementation and effectiveness of large-scale social learning.

To achieve social and academic heterogeneity for effective social learning, various grouping criteria considering individual differences have been studied (Chen & Chang, 2014). Heterogeneous grouping could be an effective method for encouraging divergent perspectives and fostering student cognitive restructuring and problem solving (Wilkinson & Fung, 2002; Zurita, Nussbaum, & Salinas, 2005). Heterogeneity promotes interactions among peers with different ability levels and backgrounds during problem solving through the exchanging of questions and explanations (Hoffman, 2002). It can also help students request clarifications, justifications, and elaborations for problem solving (Heller, Keith, & Anderson, 1991). These factors have been shown to be particularly important for low-ability students (Lou et al., 1996). Heterogeneous groups also tend to be more task-oriented (Nhan & Nhan, 2019). In addition to the use of heterogeneous groupings in terms of abilities and achievements, such groupings have been used to facilitate the understanding of the roles of ethnicity, culture, and gender in interactions and achievements (Wilkinson & Fung, 2002; Prinsen, Volman, & Terwel, 2007; Wang, 2011). However, the manner in which gender parings affect student interactions and problem-solving behaviors during social learning is still unclear.

2.2. Gender differences in learning

Gender differences have been considered as critical factors in computer-supported collaborative learning (Prinsen, Volman, & Terwel, 2007). In the context of face-to-face discussions, males tend to provide more initial suggestions than females. However, in computer-mediated communication, females tend to provide more suggestions and tend to look on communication more favorably than males (Hiltz & Johnson, 1990). However, males tend to be more confident than females in computer-supported collaborative learning environments (Joiner, Messer, Light, & Littleton, 1998).

In terms of cognition and learning, females seem to rely more on the use of memorization and rehearsing strategies, which can limit creative behaviors (Kim, 2007). In contrast, males tend to exhibit much less regulation and often question why they are studying a particular concept (Severiens & Dam, 1997). Although significant research has indicated that there are no major gender differences in terms of creativity (Kaufman, 2006), some research has indicated that males tend to perform better in terms of creative indexes (Matud, Rodríguez, & Grande, 2007; Tsai, 2013).

Gender differences have also been observed with regard to problem-solving abilities and strategies (Zhu, 2007). Males tend to be more flexible in applying problem solving strategies (Gallagher et al., 2000), while females are more likely to adhere to strategies they have already learned (Gallagher, 1998). Female interactive behaviors and problem-solving processes are also more sensitive to partner genders (Ding & Harskamp, 2006). Males tend to learn through arguments and individual activities, resulting in a more competitive atmosphere. In contract, females tend to solve problems through the use of conversation and collaboration (Croson & Gneezy, 2009). In the field of computer science, gender differences have also been observed with respect to self-efficacy, problem-solving approaches, and information processing styles (Burnett et al., 2010).

2.3. Gender pairing in collaborative learning

Some research has indicated that both genders can benefit from a collaborative learning environment. When comparing same- and mixed-gender groups, some research has indicated that both genders tend to engage less in collaborative problem solving and propose fewer solutions when placed in mixed-gender groups compared to single-gender groups (Underwood, Underwood, & Wood, 2000). Mixed-gender also tend to exhibit less verbal interaction. Males tend to be more active in dominating discourse and are typically perceived as providing better ideas than females in mixed-gender groups (Prinsen, Volman, & Terwel, 2007). Another study indicated that males in mixed-gender groups are more deliberate in terms of problem solving than they are in single-gender groups. Female-only groups have been shown to exhibit more self-disclosure and include more of their personal opinions (Savicki, Kelley, & Oesterreich, 1998). They also tend to exchange more information during problem solving (Ding & Harskamp, 2006), send more messages, and are more satisfied with social learning (Savicki, Kelly, & Lingenfelter, 1996). Female-only groups also tend to be less argumentative in the context of problem solving and general group interactions. In contrast, males tend to have discussions with less group development and communication (Savicki, Kelley, & Oesterreich, 1998).

However, previous studies have not yielded consistent results in terms of learning performance. Some research has shown that single-gender groups are more interactive when conducting computer-based tasks, but such differences do not always translate into enhanced task performance (Underwood, Underwood, & Wood, 2000; Wilkinson & Fung, 2002). Other research has indicated that females perform better in single-gender groups compared to males (Inzlicht & Ben-Zeev, 2000; Ding & Harskamp, 2006). In Whitelock et al.'s (1993) study, females and males with both similar and different views of physics achieved similar levels of performance. However, females tend to benefit more from other females with the same views. Additionally, in majority-male groups, males tend to ignore the opinions of females. Therefore, a more balanced grouping or single-gender grouping may be helpful for reducing inequities in interactions (Sanders, 2005).

3. Methodology

Facebook was considered as a social learning platform to assist in computer science learning in this study. Students were grouped into different gender pairings and asked to solve problems collaboratively through Facebook. Student problem-solving processes were analyzed in conjunction with their learning achievement. Student attitudes toward cross-school social learning were also analyzed to understand the effects of gender pairings. Figure 1 illustrates our research model.

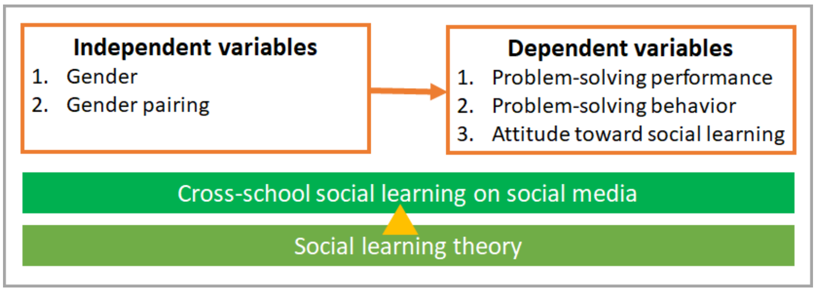


Figure 1. Research model

3.1. Participants and pairing groups

Our study participants consisted of 222 tenth-grade students (average age of sixteen) from six classes in three different high schools. School A was an all-female school, school B was a mixed-gender school, and school C was an all-male school. Students from all three schools were considered as average to high achievers based on their high school entrance exam scores. The gender pairing groups are detailed in Table 1 along with the assigned teacher for each class.

Table 1. The gender pairings for all lecture classes

Gender pairing	Lecture class			# of students
	School A	School B	School C	
Mixed-gender	A.1 (43 girls, teacher 1)	B.1 (33 boys, teacher 2)		76
Female-female	A.2 (42 girls, teacher 1)	B.2 (30 girls, teacher 2)		72
Male-male		B.3 (42 boys, teacher 2)	C.1 (32 boys, teacher 1)	74

During each lecture, each teacher explained the assigned topic to the students in their class. After each lecture, students were divided into different groups to engage in collaborative learning based on discussion through Facebook. Gender pairing groups were constructed using two lecture classes from two schools (a pool of approximately eighty students). From this pool, students were grouped into four teams consisting of approximately twenty students. Each team consisted of approximately ten students from one school and ten students from the other school. Previous studies have suggested that a suitable group size for cooperative learning ranges from two to six individuals in a face-to-face setting (Siavin, 1991) with more individuals for online settings (Hou & Wu, 2011). Because our research focused on a cross-school context and the target problem-solving tasks were relatively challenging, additional group members could provide additional social support for collaboration.

3.2. Procedure

The lecture classes were designed to teach computer science problem solving using the concepts of ***data representation*** (e.g., binary representation and data encoding/compression). All lectures were presented by two teachers who used identical teaching materials (Figure 2). Prior to the experiment, the details of the lectures and data collection were explained by the teachers and the students who participated in the study signed consent forms. Three private groups were set up on Facebook for the three types of gender parings. To help the students construct data representation knowledge and develop problem solving skills, we provided them with the opportunity to discuss problems on Facebook both during and after class. The experimental procedure is illustrated in Figure 3.

Figure 2. Teaching materials

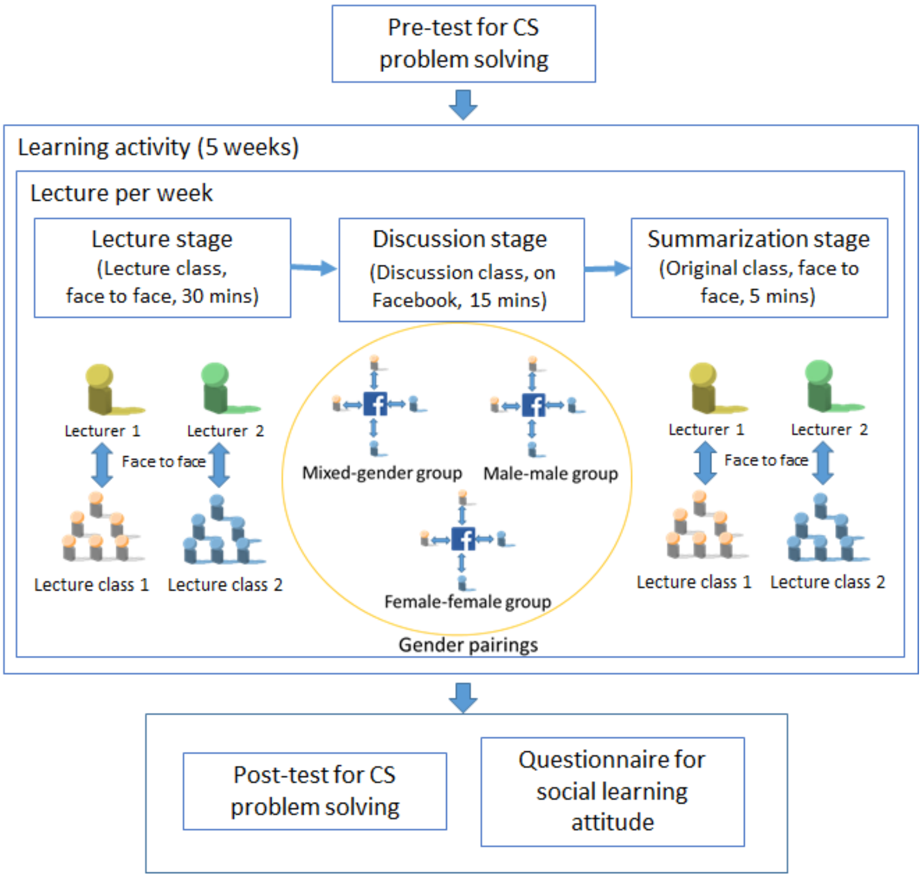


Figure 3. Experimental procedure

To analyze the initial abilities of the students, a pre-test focusing on computer science problem solving was administered prior to our experiment. Five weeks of computer science lectures were provided to the students. The two classes with same-gender pairings (e.g., A.1 and B.1, see Table 1) received the lectures during the same lecture time slots. Each of the four teams in a gender pairing class (e.g., A.1) was assigned a cooperative team from another class in the same gender pairing (B.1). During every fifty-minute in-class lecture, three instructional stages were carried out to facilitate social learning: the lecture stage, the discussion stage, and the summarization stage. In the lecture stage, the teachers explained the learning content to the students in each of their lecture classes (e.g., A.1 and B.1). In the discussion stage, teachers presented several problems to students, and asked the students to discuss solutions to the problems with their team members (e.g., a twenty-member team in the mixed-gender class) on Facebook. Students could post new ideas and respond to posts from other students in an attempt to develop collaborative solutions. The teachers did not intervene during this stage. In the summarization stage, the teachers provided a summary of what the students had discussed, as well as some additional suggestions for learning.

3.3. Instruments

3.3.1. Learning achievement test

Individual student initial abilities and learning achievements related to computer science problem solving were evaluated through a pre-test and post-test. The pre-test problems consisted of binary translation puzzles with thirteen blanks (see http://cse4k12.org/). Students had to solve the puzzle by applying their prior knowledge regarding binary translation.

Learning achievement was measured by examining two types of abilities: (1) the ability to apply what the students had learned to solve problems and (2) creativity in problem solving. Scores were graded according to the correctness of each problem-solving step. The item discriminations (point-biserial correlation coefficients) for application and creativity were 0.700 and 0.693, respectively. Q1 and Q2 represent the sample items for application (applying run-length coding) and creativity (developing a new algorithm for encryption) tests, respectively.

Q1. The vocabulary has nine words: {a, good, bad, excuse, is, worse, than, none, that}. Statistically, the frequencies of these words are 44, 32, 31, 17, 50, 28, 14, 19, 25, respectively. Please encode these words into binary codes and determine the code for the sentence "a bad excuse is worse than none."

Q2. The following paragraph appears in the article "A Grain of Sand" by William Blake. Please develop a method to encrypt this paragraph. You can use any symbol for representation (e.g., 0 or 1, black or white), but the code length should be as short as possible.
"To see a world in a grain of sand,
and a heaven in a wild flower,
hold infinity in the palm of your hand,
and eternity in an hour."

3.3.2. Collaborative problem solving behaviors

To understand the collaborative problem-solving processes used by the students, their discussion content (including posts and replies) were recorded and analyzed using a quantitative content analysis method (Pena-Shaffa & Nicholls, 2004). The discussion content was first encoded by a modified model developed from Wallas's (1926) problem solving stages, Garrison's et al. (2001) critical inquiry model, and Hou and Wu's (2011) knowledge construction and social interaction model. This modified model for collaborative problem solving encodes discussion content into six types (Table 2). The discussion types are mapped to six codes (C1 to C6) with twelve indices for further analysis.

Table 2. The coding table for discussion content

Type	Code	Index	Description
1. Problem identification	C11	Problem statement	Recognize and define the problem, and propose questions to understand the problem definition
	C12	Problem analysis	Analyze the problem elements and structure
2. Exploration	C21	Agreement	Agree with others' opinions
	C22	Disputation	Dispute others' opinions
	C23	Information providing	Provide information or cues for discussion
3. Integration	C31	Idea aggregation	Aggregate various opinions
	C32	Concept approval	Support opinions with existing literature or theories
4. Solution generation	C41	Solution testing	Verify and revise current solutions
	C42	Solution formation	Generate the final solution
5. Social interaction	C51	Social communication	Social interaction relevant to the learning task
	C52	Task coordination	Instructions or comments about task coordination
6. Digression	C61	Off-topic discussion	Posts that are completely irrelevant to the learning task

3.3.3. Attitudes toward social learning

Student attitudes toward cross-school social learning were evaluated using a questionnaire after all learning stages were completed. The questionnaire consisted of six Likert-type questions (see Table 3) with a Cronbach's alpha value of 0.87.

Table 3. The questionnaire questions

Question no.	Question
Q1	I like to use social media to assist collaborative learning.
Q2	Cross-school discussion on social media helped me improve the discussion quality in class.
Q3	Social media reduced my stress during collaborative learning.
Q4	Cross-school discussion on social media enhanced my learning.
Q5	Cross-school discussion on social media helped me express my ideas during discussion.
Q6	Social media engaged me in collaborative learning.

3.4. Data collection and analysis

ANCOVA tests were conducted to examine the effects of gender, gender pairings, and their interactions to determine whether these factors influence student learning. The pre-test grades were considered as the covariance for ANCOVA to account for the effects of the initial problem solving abilities of different groups.

After coding all posts and replies on Facebook according to the coding table (Table 2), the encoded discussion sequences for different gender pairings were analyzed using a sequential analysis method to identify significant discussion sequences.

4. Results and discussion

4.1. Learning achievement for different gender pairings

Descriptive statistics for the pre-test and post-test scores, and the ANCOVA results, are presented in Tables 4 and 5, respectively. The scores for both application and creativity range from 0 to 50.

Table 4. The descriptive statistics of the pre-test and post-test scores of problem-solving abilities for different gender pairing groups

Gender	Gender pairing	*N*	Pre-test grade		Post-test: Application		Post-test: Creativity	
			Mean	*SD*	Mean	*SD*	Mean	*SD*
Male	Mixed gender	33	9.88	3.92	23.33	15.75	37.88	18.88
	Single gender	74	10.93	3.82	28.65	16.33	37.16	21.20
	Total	107	10.61	3.86	27.01	16.27	37.38	20.42
Female	Mixed gender	43	11.16	3.68	17.44	13.82	24.42	21.47
	Single gender	72	10.46	4.12	21.67	15.56	30.21	23.34
	Total	115	10.72	3.96	20.09	15.01	28.04	22.74
Total	Mixed gender	76	10.61	3.81	20.00	14.88	30.26	21.34
	Single gender	146	10.70	3.96	25.21	16.28	33.73	22.47
	Total	222	10.67	3.90	23.42	15.97	32.55	22.10

Table 5. The ANCOVA results of problem-solving performance for genders, gender pairings, and their interactions

Performance	Factor	*F*	*p*	Cohen's *d*
Application	Gender	8.666	$.004^{**}$	0.40
	Gender pairing	4.577	$.034^{*}$	0.29
	Gender × gender pairing	0.027	.871	0.00
Creativity	Gender	11.508	$.001^{**}$	0.46
	Gender pairing	0.630	.428	0.11
	Gender × gender pairing	1.471	.227	0.17

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

In terms of gender pairings, the single-gender groups exhibited better performance than the mixed-gender groups in terms of applying their knowledge to problem solving ($F_{(1,220)} = 4.577$, $p < .05$). Interactive effects between genders and gender pairing groups cannot be observed. One can see that significant differences exist in terms of the problem solving performances between genders. These differences can be observed in terms of both application ($F_{(1,220)} = 8.666$, $p < .01$) and creativity ($F_{(1,220)} = 11.508$, $p < .01$). Male students tended to exhibit better application and creativity performance (adjusted mean: application 26.05, creativity 37.67) than female students (adjusted mean: application 19.52, creativity 27.23).

Our results indicate that single-gender groups outperformed mixed-gender groups with regard to the ability to apply computer science knowledge to problem solving. This finding differs from the results of some previous studies indicating that only females performed better in single-gender groups when compared to mixed-gender groups (Keogh, Barnes, Joiner, & Littleton, 2000; Ding, Bosker, & Harskamp, 2011). Other research by Light et al. (2000) indicated that male performance was also better when working with only males. This might because males were often eager to dominate discussions in mixed-gender groups or because including females in a discussion led to more divergent opinions. In some cases, females tended to become argumentative with respect to male opinions, which sometimes manifested as shallow thinking. In mixed-gender groups, there was sometimes a tendency for males to spend large amounts of time presenting information and ideas to group members, perhaps in an attempt to display their abilities to the females in the group. This often prevented the groups from progressing further and developing their solutions. In some mixed-gender groups, a significant amount of time was spent proposing diverse opinions that could not be integrated to find clear solutions to problems. In contrast, both males and females in single-gender groups tended to have highly focused discussions. Additionally, male-male groups tended to cooperate mainly with peers from the same school. They were highly motivated to find solutions to problems to outperform the students from other schools. Some previous research has also indicated that males are often motivated to support their own groups, particularly when faced with threats from other groups (Matud, Rodríguez, & Grande, 2007).

In this study, learning activities were done across school boundaries in a social setting. The fact that students were mixed together in online discussions helped promote more divergent discussions in mixed-gender groups, which differs from the results of many previous studies. Single-gender groupings have previously tended to promote more focused discussion because members were not distracted by competitive conversations with the opposite sex. Prior research has also shown that single-gender groups often paid more attention to discussion tasks, resulting in higher grades in the application phase of testing.

4.2. Collaborative problem solving behaviors for different gender pairings

The extracted discussion content (3257 messages) was encoded according to the codes in Table 2. The coded results regarding the proportions of various discussion types are presented in Tables 6 and 7. All gender pairing groups posted more than 1000 messages. Most discussions were aimed at exploring information and ideas (43.2), while a small number of discussions focused on the integration of ideas (1.1%). The female-female class spent much more time on idea integration (25%) than the other two classes (mixed: 8%, male-male: 2%). The females spent less time on off-topic discussion (11.8%) than the other classes (mixed: 17.0%, male-male: 23.8%). The male-male class spent less time on problem identification (1.3%) than the other classes (mixed: 5.1%, female-female: 5.5%). The male-male classes spent more time working on solution generation (4.7%) than the other gender pairing groups. They spent some time on solution testing (3.3%), but spent much more time on digressive chat (23.8%) than the other pairings. The mixed-gender group spent more time on disputing the opinions of others (C22: 9.4%) than the other two groups (male-male: 5.9%, 5.6%). The female-female group tended to integrate opinions via aggregation (C31: 2.3%), but the other two groups seldom integrated their opinions (mixed: 0.7%, male-male: 0.2%).

Table 6. The number and proportion of messages posted by different gender pairing groups for each discussion type

Gender pairing	Discussion type						Total
	Problem identification	Exploration	Integration	Solution generation	Social interaction	Digression	
Mixed	58(5.1%)	476(42.2%)	8(0.7%)	22(2.0%)	371(32.9%)	192(17.0%)	1127
Male-male	14(1.3%)	447(43.0%)	2(0.1%)	49(.7%)	281(27.0%)	248(23.8%)	1041
Female-female	60(5.5%)	482(44.3%)	25(2.3%)	33(3.0%)	361(33.1%)	128(11.8%)	1089
Average	44(4.1%)	468(43.2%)	12(1.1%)	35(3.2%)	338(31.1%)	189(17.4%)	1086

The encoded messages were than analyzed using sequential analysis to elucidate the collaborative problem solving processes exhibited during social learning on Facebook. Figure 4 presents the resulting problem-solving processes for all gender pairings. Arrows are used to indicate significant sequences in discussion content and the numbers on the arrows represent *Z*-scores. In the mixed-gender class, students tended to discuss computer science problems by providing supportive suggestions. During discussion, they also exhibited significant social communication about the learning topic or discussed other off-topic subjects. In the male-male class, students exhibited similar collaborative problem solving processes. However, the male-male class conducted additional focused discussions to test their possible solutions. In the female-female class, students tended to identify and analyze problems before they explored solutions. There was also a large amount of social communication during the information exploration stage and the students often returned to the problem identification stage following this social communication.

Table 7. The proportion (%) of messages posted by different gender pairing groups for each discussion code

Gender pairing	Discussion type											
	C11	C12	C21	C22	C23	C31	C32	C41	C42	C51	C52	C61
Mixed	2.4	2.8	2.1	9.4	30.7	0.7	0.0	1.1	0.9	31.1	1.9	17.0
Male-male	0.4	1.0	3.0	5.9	34.1	0.2	0.0	3.3	1.4	25.3	1.7	23.8
Female-female	2.7	2.8	1.5	5.6	37.2	2.3	0.0	1.5	1.6	32.0	1.2	11.8

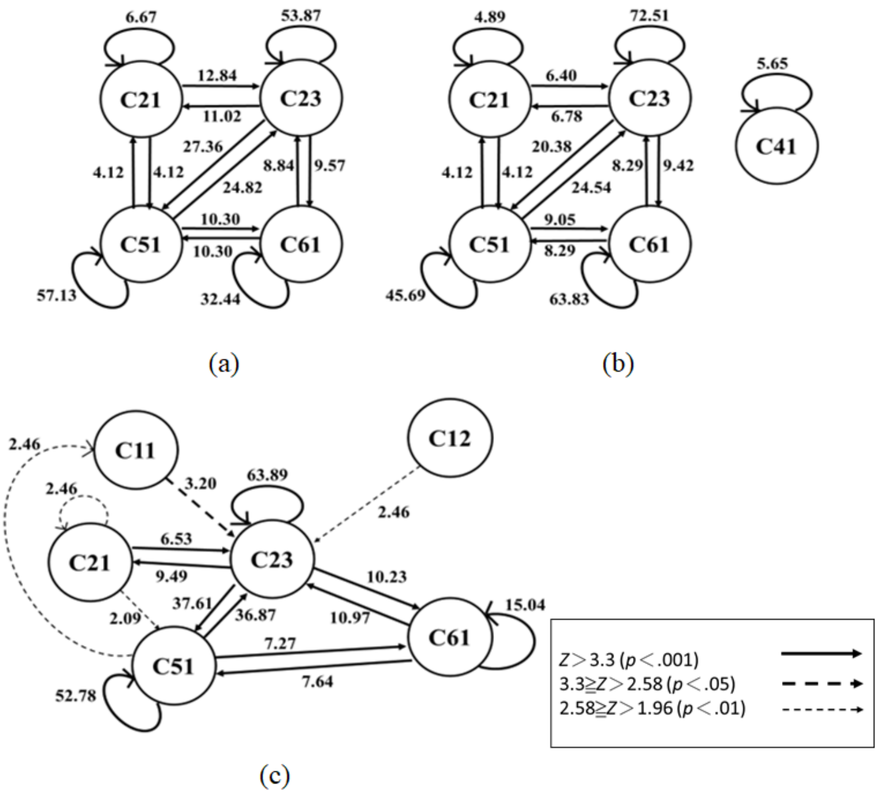


Figure 4. The collaborative problem solving processes during social learning for (a) the mixed-gender, (b) the male-male, and (c) the female-female groups

It should be noted that two types of interactions in the mixed-gender group were identified: male-dominated discussions in which males quickly proposed solutions to a problem before females could propose their solutions and very interactive discussions involving both males and females. In the first case, males often proposed solutions to problems without even joining the discussion series posted by the females. In the first type of interaction, males tended to engage in frequent disputes with other males in the mixed-gender group, apparently eager to demonstrate their abilities to the females in the group. Consequently, the mixed-gender group made many more posts (approximately twice as many) than the single-gender groups. Keogh et al. (2000) found similar results, where males in mixed-gender groups tended to dominate discussions and had many more disagreements than those in single-gender groups.

One of the problems for discussion focused on Huffman coding.

In the dictionary, we have the words: I, you, love, hate, and, the, spider, flower.
The corresponding frequencies of the words are: 18, 14, 6, 3, 20, 32, 2, 4.
Please encode the sentence: I love you and hate the spider.

The following conversation from a mixed-gender group shows how males tended to dominate the group discussion. Students S1, S2, and S3 are all males. The males in this group also attempted to find errors in their solutions to the problems and voiced differing opinions (disagreements) related to problem solutions.

S1: The code is 0111011101111101011111 1.

S2: 01 110 1110 111110 00 10 111111, I love and hate the spider

S1: I think something should be added to the front and rear.

S3: No. It is not necessary to add those elements for Huffman coding. You are thinking of Morse code.

S1: S2, you missed a word: YOU!

In the second type of interaction in mixed-gender groups, males and females often had diverse perspectives and had difficulty coordinating their discussions. Consequently, they could not agree on a clear direction for solving their computer science problems. The following conversation on Facebook is a good example of how such conversations proceeded in the mixed-gender group. Students S4, S5, and S8 are males, while students S6, S7, and S9 are females. This group was examining a problem involving Huffman coding. They provided various ideas and information about Huffman coding, but after a long and non-focused discussion, they could not identify an appropriate direction for developing solutions to the Huffman tree.

S4: Let me draw the coding tree!

S5: I have no idea.

S6: S4, have you finished that?

…

S7: Let's discuss the pros and cons of this coding method.

S6: Ok.

S7: The pros: it can be used for secret communication.

S7: It is not easy to crack, but is easy to compress.

S8: We need more people involved in this discussion.

S6: The cons: it is not suitable for text content.

S5: It is simple RAR compression.

S6: There will be a lot of numbers.

…

S9: 1101111011101011111100111111 1

S6: I don't understand.

S5: WTFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF

…

However, in many cases, the male-male group spent little time on information exploration and often proceeded to propose and test solutions to problems directly (Figure 4(b)). High-achievement students often quickly found strategies to solve problems and solicited other students to support them. Compared to the other two groups, the male-male group tended to solve problems more effectively. The following example illustrates the effect of having high-achievement students in a student group. S10 and S11 are high-achievement students (they achieved perfect scores on both the pre-test and post-test). These two students proposed the initial steps for the problem solution by listing the frequency of each word and drawing the corresponding Huffman tree (Figure 5). S14 from their group then agreed with their ideas and assigned each word a code based on the frequencies and Huffman tree. Therefore, some hints in the initial stage may motivate successful ideas for problem solving.

S10: I, you, love, hate, and, the, spider, flower. The frequencies: 18, 14, 6, 3, 20, 32, 2, 4.

S11: I drew a Huffman tree like this.

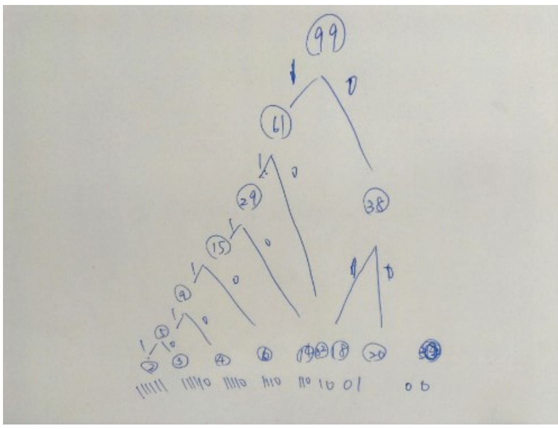


Figure 5. The Huffman tree drawn by the student

S10: I→(0,1) love→(1,1,1,0)
S12: I don't understand.
S10: you→(1,1,0) and→(0,0)
S13: Good afternoon.
S14: the→(1,0) and→(0,0) flower→(1,1,1,1,0) hate→(1,1,1,1,1,0)
S15: The advantages and disadvantages of the Huffman tree: this algorithm compresses the text or binary files efficiently…
S14: spider→(1,1,1,1,1,1,)
S16: http://www.delightpress.com.tw/bookRead/sknd00004_read.pdf
S15: The main feature of Huffman coding: the resulting code should be unique…

4.3. Attitudes toward social learning for different gender pairings

Student attitudes toward social learning on Facebook were obtained from six four-point Likert-type scale questions in a post-experiment questionnaire (1: strongly disagree, 2: disagree, 3: agree, 4: strongly agree). Table 8 presents descriptive statistics for the social learning attitudes of different genders and gender pairings, and Table 9 presents analysis of variance (ANOVA) results for genders, gender pairings, and their interactions. Because a Levene test showed the variances to be significantly different between genders, as well as between the interactions between genders and gender pairings, a Welch test was used instead of a traditional *F* test to analyze these differences. The results indicate that females agree with the effectiveness of social learning more than males ($F_{(1,220)} = 9.715$, $p < .01$). Additionally, the genders and gender pairings exhibit interactive effects ($F_{(2,219)} = 4.085$, $p < .01$). For example, students in the female-female group had much more positive attitudes than those in the male-male group (Table 10). However, students in the mixed- and single-gender groups had similar perceptions regarding social learning.

Table 8. Descriptive statistics regarding attitudes toward social learning

Gender	Gender pairing	*N*	Questionnaire result	
			Mean	*SD*
Male	Mixed	33	2.68	0.64
	Single	74	2.54	0.68
	Total	107	2.58	0.67
Female	Mixed	43	2.77	0.57
	Single	72	2.86	0.44
	Total	115	2.83	0.49
Total	Mixed	76	2.73	0.60
	Single	146	2.70	0.59
	Total	222	2.71	0.59

Table 9. The ANOVA results of attitudes toward social learning

Factor	*F*	*p*	Cohen's *d*
Gender	Welch: 9.715	$.002^{**}$	0.42
Gender pairing	0.179	.673	0.00
Gender×gender pairing	Welch: 4.085	$.008^{**}$	0.47

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

Table 10. Pairwise comparisons of attitudes toward social learning for different gender pairing groups

Group (I)	Group (J)	Mean difference (I-J)	*p*
Male-mixed	Male-single	0.147	.702
	Female-mixed	-0.087	.925
	Female-single	-0.179	.467
Male-single	Female-mixed	-0.234	.195
	Female-single	-0.326	$.004^{**}$
Female-mixed	Female-mixed	-0.092	.800

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

The statistical results indicate that females in the female-female groups had better attitudes toward social learning than males in the male-male groups. The responses to the open-ended questions reveal that males in the male-male group wanted to interact with and learn from females. This could explain why some males in the male-male groups exhibited more negative attitudes. Previous research has indicated that males often use social media to form new relationships (Muscanell & Guadagno, 2012). Therefore, some males might have wished to form relationships with females from other schools, as indicated by the following feedback:

"We should discuss the problems with the girl's high school."

"I am strongly opposed to discussion with another boy's school. We should discuss the problems with girls to understand diverse thinking methods."

Males in the mixed-gender groups were more positive about using social media, perhaps because females were included in the discussion. Males in the mixed-gender groups were often eager to discuss computer science problems, perhaps to impress the females. This could explain why the mixed-gender groups tended to have more disagreements with respect to problem solving because males might have been competing for attention from the females in the group. Some males preferred to interact more with females, perhaps based on the belief that more diverse discussions would occur, as indicated by the following feedback:

"It was helpful to discuss the problems with students from another school, especially the girls. This made me more engaged in the discussion."

"This was a great learning activity and I like to learn with girls."

Although some students in the male-male group initially disliked working with males from another school, most group members were still able to consider the ideas of others and benefit from the divergence of discussions. A few students provided the following comments:

"I was exposed to various ideas, which fostered my multi-perspective thinking."

"I could observe and learn from various opinions."

5. Implications for instructional practices

The insights gleaned from this study can aid educators in the design of more adaptive teaching/grouping strategies, particularly in the area of problem solving. Single-gender grouping helps students focus on problem solving and allow students to apply the information they know to solving problems more efficiently. However, the diversity of discussion induced by mixed-gender groupings is often necessary to facilitate breakthroughs with regard to expanding existing thinking and reaching solutions to complex problems. More efficient instructional guidance should be employed to improve student enthusiasm and provide opportunities to listen to diverse ideas. Diverse ideas are a critical component in the process of creative thinking (Jang, 2009). Our findings provide useful insights for applying gender pairing to foster effective group discussions. In the initial stages of problem solving, it appears that mixed-gender groups may be more effective at inspiring more diverse opinions and creative ideas. However, in the later stages of problem solving, it appears that single-gender groups may be more effective at helping students reach convergent solutions.

In single-gender schools, cross-school collaborative learning through social media may introduce new and divergent ideas to students. In mixed-gender schools, collaborative learning with single-gender groupings may help foster more focused and deep discussions among students. We also found that cross-school competition encourages students to be more active in problem solving, particularly males. Although some researchers have argued that online learning eliminates the gender cues and encourages more participation and cooperation (Arbaugh, 2000), a cross-school setting may stimulate competition and encourage more participation. Appropriate instructional strategies (e.g., peer review) could be designed to help students benefit from competition and cooperation in social learning.

To extend the applicability of our research findings to other problem-solving scenarios, it should be noted that instructional strategies should be adapted according to student ages, learning subjects, and social media platforms to improve effectiveness.

6. Conclusions

This study addressed three research questions and empirical analysis was conducted to explore the effects of gender pairings on collaborative problem-solving performance, processes, and attitudes in a social learning environment. Our findings for each of the three research questions can be summarized as follows. (1) Single-gender groups exhibited better performance than mixed-gender groups in terms of applying acquired knowledge to problem solving. Additionally, significant differences existed in terms of problem solving behaviors between genders. Males tended to exhibit higher levels of both application and creativity performance than females. (2) Students in male-male groups tended to focus on problem solving, although they also engage in more off-topic discussion. They also tended to develop and test solutions directly without spending much time on problem identification. In mixed-gender groups, males tended to dominate the discussion and different genders tended to propose divergent opinions. (3) Females exhibited better attitudes toward social learning activities. This was particularly evident in the female-female groups, which could be attributed to the fact that students could observe many diverse ideas. Additionally, although the male-male groups exhibited the highest levels of performance, they also expressed more negative attitudes toward an interactive environment based on a lack of female participation. The findings and implications of this study can facilitate additional research on the use of social learning to develop effective instructional strategies.

Acknowledgement

We thank the Ministry of Science and Technology of Taiwan for funding this study (MOST 106-2511-S-003-031-MY3, MOST 109-2511-H-003-010-MY3, MOST 109-2511-H-003-009-), and Dr. Fang-Lan Huang for the assistance in the instructional experiment.

References

Al-Rahmi, W. M., Alias, N., Othman, M. S., Marin, V. I., & Tur, G. (2018). A Model of factors affecting learning performance through the use of social media in Malaysian higher education. *Computers & Education*, *121*, 59-72.

Al-Zahrani, A. M. (2015). From passive to active: The impact of the flipped classroom through social learning platforms on higher education students' creative thinking. *British Journal of Educational Technology*, *46*(6), 1133–1148.

Arbaugh, J. B. (2000) Virtual classroom versus physical classroom: An Exploratory comparison of class discussion patterns and student learning in an asynchronous internet based MBA course. *Journal of Management Education, 24*, 207–227.

Avcı Yücel, Ü., & Usluel, Y. K. (2016). Knowledge building and the quantity, content and quality of the interaction and participation of students in an online collaborative learning environment, *Computers & Education*, *97*, 31–48.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

Berkes, F. (2009). Evolution of co-management: Role of knowledge generation, bridging organizations and social learning. *Journal of Environmental Management, 90*(5), 1692–1702.

Burnett, M., Fleming, S. D., Iqbal, S., Venolia, G., Rajaram, V., Farooq, U., Grigoreanu, V., & Czerwinski, M. (2010). Gender differences and programming environments: Across programming populations. In *Proceedings of the 2010 ACM-IEEE international symposium on empirical software engineering and measurement* (pp. 1-10). doi:10.1145/1852786.1852824

Chen, C.-M., & Chang, C.-C. (2014). Mining learning social networks for cooperative learning with appropriate learning partners in a problem-based learning environment. *Interactive Learning Environments, 22*(1), 97–124.

Chung, K. S. K., & Paredes, W. C. (2015). Towards a social networks model for online learning & performance. *Educational Technology & Society*, *18*(3), 240–253.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, *47*(2), 1-27.

Ding, N., & Harskamp, E. (2006). How partner gender influences female students' problem solving in physics education. *Journal of Science Education and Technology*, *15*(5), 331-343.

Ding, N., Bosker, R. J., & Harskamp, E. G. (2011). Exploring gender and gender pairing in the knowledge elaboration processes of students using computer-supported collaborative learning. *Computers & Education*, *56*(2), 325–336.

Douglas, S., Farley, A., Lo, G., Proskurowski, A., & Young, M. (2010). Internationalization of computer science education. In *Proceedings of the 41st ACM technical symposium on Computer science education* (pp. 411-415). doi:10.1145/1734263.1734404

Eid, M. I. M., & Al-Jabri, I. M. (2016). Social networking, knowledge sharing, and student learning: The Case of university students. *Computers & Education*, *99*, 14–27.

Gallagher, A. M. (1998). Gender and antecedents of performance in mathematics testing. *Teachers College Record*, *100*, 297–314.

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A V, Morely, M., & Cahalan, C. (2000). Gender differences in advanced mathematical problem solving. *Journal of Experimental Child Psychology*, *75*(3), 165–190. doi:10.1006/jecp.1999.2532

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, *15*(1), 7–23.

Heller, P., Keith, R., & Anderson, S. (1992). Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving. *American Journal of Physics, 60*(7), 627-636.

Heo, H., Lim, K., & Kim, Y. (2010). Exploratory study on the patterns of online interaction and knowledge co-construction in project-based learning. *Computers & Education*, *55*(3), 1383-1392.

Hiltz, R. S., & Johnson, K. (1990). User satisfaction with computer-mediated communication systems. *Management Science*, *36,* 739-751.

Hoffman, J. (2002). Flexible grouping strategies in the multiage classroom. *Theory into practice, 41*(1), 47-52.

Hou, H. T., & Wu, S. Y. (2011). Analyzing the social knowledge construction behavioural patterns of an online synchronous collaborative discussion instructional activity using an instant messaging tool: A Case study. *Computers & Education*, *57,* 1459-1468.

Inzlicht, M., & Ben-Zeev, T. (2000). A Threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, *11*, 365–371.

Jang, S.-J. (2009). Exploration of secondary students' creativity by integrating web-based technology into an innovative science curriculum. *Computers & Education, 52*(1), 247–255.

Joiner, R., Messer, D., Light, P. & Littleton, K. (1998). The Effects of gender, expectations of success and social comparison on children's performance on a computer-based task. *Educational Psychology*, *18*(3), 319-326.

Keogh, T., Barnes, P., Joiner, R., & Littleton, K. (2000). Computers, verses, paper–girls versus boys: Gender and task presentation effects. *Educational Psychology*, *20*(1), 33-44.

Kim, K. H. (2007). Exploring the interactions between Asian culture (Confucianism) and creativity. *The Journal of Creative Behavior*, *41*(1), 28–53.

Kimmerle, J., Moskaliuk, J., & Cress, U. (2011). Using Wikis for learning and knowledge building: Results of an experimental study. *Educational Technology & Society*, *14*(4)*,* 138–148.

Light, P., Littleton, K., Bale, S., Joiner, R., & Messer, D. (2000). Gender and social comparison effects in computer-based problem solving. *Learning and Instruction*, *10*(6), 483–496.

Lin, P. C., Hou, H. T., Wang, S. M., & Chang, K. E. (2013). Analyzing knowledge dimensions and cognitive process of a project-based online discussion instructional activity using Facebook in an adult and continuing education course. *Computers & Education*, *60*(1), 110-121.

Liu, C. C., & Tsai, C. C. (2008). An Analysis of peer interaction patterns as discoursed by on-line small group problem-solving activity. *Computers & Education*, *50*, 627-639.

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A Meta-analysis. *Review of Educational Research, 66*(4), 423–458.

Matud, M. P., Rodríguez, C., & Grande, J. (2007). Gender differences in creative thinking. *Personality and Individual Differences*, *43*(5), 1137–1147.

Moghavvemi, S., Sulaiman, A., Jaafar, N. I., & Kasem, N. (2018). Social media as a complementary learning tool for teaching and learning: The case of youtube. The *International Journal of Management Education*, *16*(1), 37-42.

Moreno, L., Gonzalez, C., Castilla, I. Gonzalez, E., & Sigut, J. (2007). Applying a constructivist and collaborative methodological approach in engineering education. *Computers & Education*, *49*, 891-915.

Moskaliuk, J., Kimmerle, J., & Cress, U. (2012). Collaborative knowledge building with wikis: The Impact of redundancy and polarity. *Computers & Education*, *58*(4), 1049-1057.

Muscanell, N. L., & Guadagno, R. E. (2012). Make new friends or keep the old: Gender and personality differences in social networking use. *Computers in Human Behavior*, *28*(1), 107–112.

Nhan, H., & Nhan, T. A. (2019). Different grouping strategies for cooperative learning in English majored seniors and juniors at Can Tho University, Vietnam. *Education Sciences, 9*(1), 59. doi:10.3390/educsci9010059

Pena-Shaff, J. B., & Nicholls, C. (2004). Analyzing student interactions and meaning construction in computer bulletin board discussions. *Computers & Education*, *42*(3), 243–265.

Prestridge, S., Tondeur, J., & Ottenbreit-Leftwich, A. T. (2019). Insights from ICT-expert teachers about the design of educational practice: The Learning opportunities of social media. *Technology, Pedagogy and Education*, *28*(2), 157-172.

Prinsen, F. R., Volman, M. L. L., & Terwel, J. (2007). Gender-related differences in computer-mediated communication and computer-supported collaborative learning: Original article. *Journal of Computer Assisted Learning*, *23*(5), 393–409.

Puntambekar, S. (2006). Analyzing collaborative interactions: Divergence, shared understanding and construction of knowledge. *Computers and Education*, *47*(3), 332-351.

Ranieri, M., Manca, S., & Fini, A. (2012). Why (and how) do teachers engage in social networks? An Exploratory study of professional use of Facebook and its implications for lifelong learning. *British Journal of Educational Technology, 43*(5), 754-769.

Sanders, J. (2005). *Gender and technology in education: A Research review*. Retrieved from www.josanders.com/pdf/gendertech0705.pdf

Savicki, V., Kelley, M., & Lingenfelter, D. (1996). Gender, small group composition, and task type in small task groups using computer-mediated communication. *Computers in Human Behavior*, *12*(4), 549-565.

Savicki, V., Kelley, M., & Oesterreich, E. (1998). Effects of instructions on computer-mediated communication in single- or mixed-gender small task groups. *Computers in Human Behavior, 14*, 163-180.

Severiens, S., & Dam, G. T. (1997). Gender and gender identity difference in learning styles. *Educational Psychology, 17*, 79-93.

Siavin, R. E. (1991). Synthesis of research on cooperative learning. *Educational Leadership*, *48*(5), 71–82.

Son, J., Kim, J. D., Na, H. S., & Baik, D. K. (2016). A Social Learning Management System supporting feedback for incorrect answers based on social network services. *Educational Technology and Society*, *19*(2), 245–257.

Tsai, K. C. (2013). Examining gender differences in creativity. *The International Journal of Social Sciences, 13*(1), 115-122.

Underwood, J., Underwood, G., & Wood, D. (2000). When does gender matter? Interactions during computer-based problem solving. *Learning and Instruction*, *10*(5), 447-462.

Vygotsky, L. (1978). *Mind in society*. London, UK: Harvard University Press.

Wallas, G. (1926). *The Art of thought.* New York, NY: Harcourt, Brace.

Wang, C. M. (2011). Instructional design for cross-cultural online collaboration: Grouping strategies and assignment design. *Australasian Journal of Educational Technology, 27*(2), 243-258.

Whitelock, D., O'Shea, T., Taylor, J., Scanlon, E., Clark, P., & O'Malley, C. (1993). The Interaction between group composition and computer supported collaborative learning. *Paper presented at the Computers in Psychology Conference*, York.

Wilkinson, I. A. G., & Fung, I. Y. Y. (2002). Small-group composition and peer effects. *International Journal of Educational Research*, *37*(5), 425–447. doi:10.1016/S0883-0355(03)00014-4

Wong, L. H., & Hsu, C. K. (2016). Effects of learning styles on learners' collaborative patterns in a mobile-assisted, Chinese character-forming game based on a flexible grouping approach. *Technology, Pedagogy and Education*, *25*(1), 61-77.

Zhan, Z., Fong, P. S., Mei, H., & Liang, T. (2015). Effects of gender grouping on students' group performance, individual achievements and attitudes in computer-supported collaborative learning. *Computers in Human Behavior*, *48*, 587-596.

Zhu, Z. (2007). Gender differences in mathematical problem solving patterns: A Review of literature. *International Education Journal*, *8*(2), 187-203.

Zurita, G., Nussbaum, M., & Salinas, R. (2005). Dynamic grouping in collaborative learning supported by wireless handhelds. *Educational Technology and Society*, *8*(3), 149–161.

Yeh, Y., Chang, C.-Y., Ting, Y.-S., & Chen, S.-Y. (2020). Effects of Mindful Learning Using a Smartphone Lens in Everyday Life and Beliefs toward Mobile-based Learning on Creativity Enhancement. *Educational Technology & Society*, *23* (4), 45–58.

Effects of Mindful Learning Using a Smartphone Lens in Everyday Life and Beliefs toward Mobile-based Learning on Creativity Enhancement

Yu-chu Yeh[1,2*], Chih-Yen Chang[3], Yu-Shan Ting[4] and Szu-Yu Chen[5]

[1]Institute of Teacher Education, National Chengchi University, Taiwan // [2]Research Center for Mind, Brain & Learning, National Chengchi University, Taiwan // [3]Department of Physical Education, National Taiwan Normal University, Taiwan // [4]Department of Education, National Chengchi University, Taiwan // [5]Department of Business Administration, Oklahoma State University, USA // ycyeh@nccu.edu.tw // eally1123@gmail.com // pnm40275@gmail.com // cindytaipeior@gmail.com

[*]Corresponding author

(Submitted June 30, 2020; Revised August 12, 2020; Accepted September 7, 2020)

ABSTRACT: To date, no study has examined whether enhancing mindfulness in everyday life through varied smartphone-based interventions involving photo taking could improve creativity or whether the beliefs toward mobile-based learning would influence such mindfulness learning effect. This study, therefore, developed the inventory "Beliefs toward Mobile Devices for Creativity Learning" and designed four interventions to validate the feasibility of such a new approach. One hundred and eighty-three college students participated in the inventory development, and 149 college students, who were randomly assigned to a control group (Group 1) or one of the three experimental groups, participated in a one-week experimental instruction. While Group 1 did not receive any smartphone intervention, the experimental groups used their smartphones to take photos with different emphases for four days and to share the photos with imaginative narratives on a designated website. Group 2 emphasized free choices of photo taking (self-determination), Group 3 emphasized self-determination and idea sharing, and Group 4 emphasized self-determination in varied categories and idea sharing. The results suggest the developed inventory has good reliability and validity; moreover, incorporating both self-determination and knowledge sharing lead to the best learning effect. Notably, beliefs toward mobile-based learning influence intervention effects on the enhancement of creativity self-efficacy. Accordingly, even a small amount of mindful learning in everyday life using a smartphone lens may enhance creativity. This study provides a valid instrument and smartphone-based mindfulness approach for ubiquitous learning of creativity.

Keywords: Creativity, Knowledge sharing, Mindful learning, Self-determination, Smartphone

1. Introduction

The field of education has recently applied mindful pedagogies because of the known benefits of improved attention (Zilcha-Mano & Langer, 2016), cognitive flexibility (Haller, Bosman, Kapur, Zafonte & Langer, 2017), and creativity (Baas, Nevicka, & Velden, 2020; Langer, 2000; Wang & Liu, 2016; Yeh, Chang, & Chen, 2019a). Mindfulness is characterized by the continuous creation of new categories, an implicit awareness of multiple perspectives, and openness to new information (Langer, 2016); these characteristics are all critical to creativity, which has been regarded as a top educational goal across educational levels. Empirical findings have also revealed overlapping concepts between mindfulness and creativity (Davenport & Pagnini, 2016; Bercovitz, Pagnini, Phillips & Langer, 2017; Yeh et al., 2019a). In addition, a recent study (Yeh, Chen, Rega, & Lin, 2019b) showed that mindful learning enhanced self-efficacy in game-based learning. Accordingly, facilitating creativity and creative self-efficacy through mindful learning can be an effective approach for creativity instruction.

A convenient and efficient way to improve mindful learning and creativity is to practice such cognitive processes through devices people commonly use in everyday life. The smartphone has become an essential piece of equipment for people in modern societies; it is not only an important instrument for learning (Ashford, Giorgi, Mann, Sherritt, Ungar, & Curtis, 2020; Farley et al., 2015) but also for capturing beautiful moments in life. The vast majority of people enjoy sharing photos on social media using smartphones. When people take and share photos with smartphones, and if they could be a little bit more mindful and imaginative about the photos they take, would they be more creative (Question 1)? Moreover, research findings have suggested that self-determination and knowledge sharing may enhance mindful learning (Yeh, Yeh, Chen., 2012; Yeh et al., 2019a; Yeh & Lin, 2015) and creativity (Peterson, Aljadeff-Abergel, Eldridge, VanderWeele, & Acker, 2020; Kremer, Villamor, & Aguinis, 2019). Would smartphone-based mindfulness interventions with varied types of self-determination and knowledge sharing carry different effects on the improvement of people's creativity? (Question 2). Finally, identifying critical personal traits that may moderate intervention effects helps achieve optimal learning (Yeh & Lin, 2015). Would people's beliefs toward using mobile devices for creativity learning influence the smartphone-based mindfulness interventions (Question 3)?

To date, no study has employed smartphone-based mindfulness interventions through photo taking in everyday life to enhance the ability and self-efficacy of creativity. To answer the aforementioned questions, we developed an inventory and designed four interventions based on a theoretical framework, by which we attempted to examine what kind of mindfulness interventions might be more effective and whether people's beliefs toward using mobile devices for creativity learning would moderate such learning effect. Hopefully, a new smartphone-based approach would be proposed to advance the theories and ubiquitous learning of creativity.

2. Related work

2.1. Mindfulness

Mindfulness is considered a secularized adaptation of Eastern Buddhist convention, and it is usually defined as "paying attention in a particular way: on purpose, in the present moment, and nonjudgmentally" (Kabat-Zinn, 1994, p. 4). The Western psychological perspective of mindfulness was pioneered by Langer (1989). Conceiving mindfulness within a cognitive information-processing framework, Langer (2012) pointed out that mindfulness is easy to learn, thus making it appealing to those unwilling to meditate. According to Langer (2012), mindfulness can be increased by paying attention to novelty, trying to be flexible in evaluations and perceptions, and questioning previous points of view that have been taken for granted.

Education is one of the areas to which mindfulness has been commonly applied in recent years. When people engage in mindful learning, they avoid forming mindsets that unnecessarily confine them (Langer, 2000). Langer (2016) proposed that a mindful learning approach possesses three characteristics: "the continuous creation of new categories, openness to new information, and an implicit awareness of more than one perspective" (p. 4). Along the same line, Ben-Ari (2002) proposed that mindful movement is different from incidental movement in that it occurs as a result of active learning. In this study, we define mindfulness as a mindful learning process in which individuals actively and consciously pay attention to the things they are curious about or interested in, and further, try to bring about new meanings or original thinking from these ordinary or special things.

2.2. Mindfulness, creativity, and creative self-efficacy

The concept of self-efficacy was originally proposed by Bandura (1977) who claimed that self-efficacy refers to beliefs in one's abilities to organize and execute the actions essential for producing given outcomes; individuals with self-efficacy act with forethought, self-reactiveness, and self-reflectiveness (Bandura, 2001). As such, self-efficacy is an important predictor for behaviors pertaining to attention and motivational processes in learning (Bandura, 2012). According to Bandura (1997), self-efficacy involves efficacy expectations and outcome expectations, and these two expectations influence how much efforts and what actions an individual will take to achieve the learning outcomes (Bandura, 1997; Woolfolk, 2016). More recently, some researchers have employed the concept of self-efficacy in creativity studies and suggest that creative self-efficacy is critical to creative performance (Tierney & Farmer, 2002; Wang, Liu & Shalley, 2018) and that it involves one's intrinsic motivation to perform creative behaviors (Gong, Huang & Farh, 2009). Integrating the concept of self-efficacy and the learning of creativity, this study defines creative self-efficacy as the belief in one's ability to produce creative ideas or solutions and the confidence in achieving creative performance.

Regarding the relationship between mindfulness and creativity, it can be explained through the cognitive processes of creativity. Creative processes are greatly influenced by working memory and emotion (Augello et al., 2016; Vartanian, 2019; Yeh, Lai, & Lin, 2016; Yeh, Lai, Lin, Lin, & Sun, 2015). Working memory involves in integration, processing, and retrieval of information, as well as the maintenance and manipulation of task-relevant information to guide subsequent behavior (Autin & Croizet, 2012); it may influence creativity via attention to task-related information (Yeh, Tsai, Hsu, & Lin, 2014). Mindfulness involves the self-regulation of attention with curiosity and open-mindedness (Bishop et al., 2004) as well as creativity and self-regulation (Pang & Ruch, 2019); it also contributes to awareness, cognitive flexibility, and emotion regulation, which are critical to creativity performance (Yeh et al., 2019b). Accordingly, mindfulness may influence creativity through enhancing attention, working memory efficiency, and emotional regulation.

On the other hand, few studies have explored the relationship between mindfulness and creativity self-efficacy. However, the positive relationship between mindfulness and creative self-efficacy has been supported by a related review (Caldwell, Harrison, Adams, Quin, & Greeson, 2010) and empirical studies in varied contexts (e.g., Greason & Cashwell, 2009; Zheng & Liu, 2017; Yeh et al., 2019b). For example, Greason and Cashwell (2009) found that mindfulness was a significant predictor of counseling self-efficacy and that attention was a mediator of that relationship. Along the same line, Zheng and Liu (2017) found that abusive supervision on employee self-efficacy can be buffered by employee mindfulness. Mindfulness may bolster coping abilities that help decrease negative thoughts, which undermine self-efficacy (Gärtner, 2013). In addition, creative self-efficacy was found to be influential to creative performance (Wang, Liu & Shalley, 2018). In light of these positive relationships between mindfulness, creativity, and self-efficacy, we assumed that a smartphone-based learning with emphasis on mindfulness would enhance learners' creativity and creativity self-efficacy.

2.3. Mindful learning interventions through smartphones

Mindfulness is a natural human ability as well as a set of skills that can be fostered and developed via a regular meditation practice or specifically customized interventions (Iani, Lauriola, Cafaro, & Didonna, 2017). Many different mindfulness programs or interventions have been shown to be effective in varied areas (Creswell, 2017). For example, it was found that focused attention mindfulness meditation improved executive functioning, attentional control, and emotion regulation of older adults (Polsinelli, Kaszniak, Glisky, & Ashish, 2020). In a training program emphasizing mindfulness, compassion, and social-emotional skill training, teachers' sense of efficacy, interpersonal mindfulness in teaching, and the interpersonal reactivity measures of perspective taking and personal distress were improved (Tarrasch, Berger & Grossman, 2020). Studies of Langerian mindfulness have also shown positive effects on attention (e.g., Levy, Jennings, & Langer, 2001; Zilcha-Mano & Langer, 2016) and cognitive flexibility (e.g., Haller et al., 2017).

To date, most empirical studies on Langerian mindfulness used selected components from Langer's theory, for example, noticing distinctions, multiple perspectives, or producing novelty. These techniques, although unconventional and different from one study to another, were successful in inducing a state of Langerian mindfulness, which has shown positive effects on learning (e.g., Lawrie, Tuckey & Dollard, 2018; Miralles-Armenteros, Chiva-Gómez, Rodríguez-Sánchez & Barghouti, 2019; Ostafin & Kassman, 2012; Stewart & Bower, 2019) and creativity (e.g., Grant, Langer, Falk & Capodilupo, 2004; Langer, 2000; Wang & Liu, 2016). Accordingly, interventions of mindfulness may facilitate attention and cognitive flexibility and then enhance creativity and creative self-efficacy. To maximize the intervention effect, this study adds two components into the smartphone-based intervention: knowledge sharing and self-determination.

Study findings have suggested that knowledge/idea sharing enhanced the improvement of creativity and creative self-efficacy. Kremer et al. (2019) indicated that knowledge sharing is a critical factor that contributes to creativity and innovation. Yeh et al. (2019b) found idea sharing is an important mechanism for improving creativity and creative self-efficacy during game-based learning. Along the same line, Yeh and Lin's (2015) findings suggested that encouraging knowledge sharing through observation learning of peer assignments online contributed to college students' creativity improvement. Other studies also demonstrated that knowledge sharing was positively related to team creativity, which was mediated by absorptive capacity and knowledge integration (Hu, Erdogan, Jiang, Bauer & Liu, 2018; Men, Fong, Luo, Zhong & Huo, 2019; Sung & Choi, 2020; Zhang, Sun, Jiang, & Zhang, 2019). Knowledge sharing may facilitate creativity through self-awareness and self-reflection (Yeh et al., 2012) as well as paying attention to novelty, which are critical to mindful learning and creativity (Langer, 2012, 2016). This study therefore requested the participants to share their tasks on a designated website to enhance knowledge sharing and, further, facilitate mindful learning and creativity.

On the other hand, self-determination involves the concepts of choice, self-control, and self-management (Peterson et al., 2020). Self-determination theory (SDT) suggests that autonomous motivation comprises the intrinsic motivation and the internalized extrinsic motivation (Ryan & Deci, 2017); people have an active tendency toward growth and development, which leads to optimal functioning (Ryan & Deci, 2000). Three basic psychological needs required to reach this optimal functioning are autonomy, competence, and relatedness (Ryan & Deci, 2000; Vansteenkiste & Ryan, 2013). It has been found that self-determination influenced mastery experience of creativity through mindful learning experience (Yeh et al., 2019a); under the freedom of task choice condition, participants were more likely to show social creativity (Gu, Hu, Ngwira, Jing & Zhou, 2016). Based on these empirical findings, this study incorporated the concept of autonomy to enhance mindful learning and creativity. Autonomy refers to the feelings of volition, free choice, and psychological freedom in one's behavior (Long, Wang, Liu & Lei, 2019). In this study, we allowed free choices of tasks to boost autonomy.

2.4. Influences of beliefs toward smartphones on mindful learning interventions

According to aptitude-treatment interactions (ATIs), understanding the interaction between learners' aptitude and treatment helps create a learning environment in which the treatment matches the aptitude of the learner and, further, in which the optimal learning effect can be achieved (Yeh & Lin, 2015). The aptitude of concern in this study was the beliefs toward mobile devices for creativity learning. To measure this aptitude, we developed an inventory for this study called "Beliefs toward Mobile Devices for Creativity Learning" (BMD-CL). Creativity involves a set of dispositions, knowledge, and skills related to creativity (Yeh et al., 2019a). BMD-CL, therefore, refers to the beliefs that mobile devices contribute to the enhancement of creativity skills or strategies, the motivation and knowledge sharing of creative ideas, and the efficiency of creativity learning. When people hold positive beliefs toward using mobile devices to enhance creativity, they may conduct more self-regulated learning, which involves self-initiated reactions and learning motivation, and such processes facilitate task interest, task choice, and persistence (Zimmerman, 2011). Moreover, such positive beliefs may bring about enjoyment and positive emotions, which contribute to creative performance (Boyle, Connolly, & Hainey, 2011; Yeh & Lin, 2015).

2.5. Hypotheses of this study

While most college students frequently take photos with smartphones, few students mindfully learn from the photos they take. To date, no study has integrated smartphones and mindfulness interventions with emphases on self-determination and knowledge sharing to enhance college students' creativity, especially in a way that is relevant to students' daily lives. This study attempted to propose such a new approach to enhance college students' creativity and creative self-efficacy. Meanwhile, college students' beliefs toward using mobile devices to improve creativity was also considered; to achieve this goal, the BMD-CL was developed first. Specifically, this study aimed at examining whether enhancing mindfulness in everyday life through varied smartphone-based interventions (with self-determination and/or knowledge sharing) would have different effects on college students' improvement of creativity and creative self-efficacy, as well as whether BMD-CL would influence such a training effect. The following hypotheses were proposed.

- Hypothesis 1: Smartphone-based mindfulness interventions in everyday life would enhance college students' creativity. Specifically, practicing mindful learning in everyday life by taking photos and writing imaginative narratives would enhance college students' creativity.
- Hypothesis 2: Varied types of smartphone-based mindfulness interventions would have different effects on the improvement of college students' creativity. In other words, college students who took the intervention with emphases on both self-determination and knowledge sharing would outperform those who did not take any intervention or those who took the intervention without opportunities for self-determination or knowledge sharing.
- Hypothesis 3: Smartphone-based mindfulness interventions would have positive effects on college students' improvement in creative self-efficacy, and such effects would be influenced by BMD-CL. Specifically, college students who took any of the mindfulness interventions would have better improvement in creative self-efficacy than those who did not take any intervention. On the other hand, those who were at a higher level of BMD-CL would outperform those who were at a lower level of BMD-CL in creative self-efficacy.

3. Methods

3.1. Participants

All participants, aged from 20 to 30 years old, were recruited through an online ad posted on a campus website. Participants in the inventory development stage were 183 college students (61 males and 122 females; *M*age = 20.97; *SD*age = 1.469); they were rewarded with approximately $3 USD. Participants who joined the one-week experimental instruction were 160 college students; they were randomly and evenly distributed into four groups with gender consideration. However, some participants dropped out during the experimental period. Finally, the valid sample size was 149 (35 males and 114 females; *M*age = 21.21; *SD*age = 1.565): Group 1 (G1) = 40, Group 2 (G2) = 37, Group 3 (G3) = 34, and Group 4 (G4) = 38. The control group (G1), who only took the pretest and the posttest, were rewarded with approximately $7 USD, whereas the experimental groups (G2, G3, and G4), who took the pretest, intervention, and the posttest, were rewarded with approximately $35 USD. This study was approved by the Institutional Review Board of National Chengchi University, Taiwan and written informed consent was obtained from all participants.

3.2. Instruments

3.2.1. Beliefs toward Mobile Devices for Creativity Learning (BMD-CL)

The instrument BMD-CL, with 13 items, was developed to measure college students' beliefs toward mobile devices for creativity learning. BMD-CL is a 6-point Likert-type scale from 1 point to 6 points, representing strongly disagree to strongly agree. The reliability and validity information are shown in the results section.

3.2.2. Inventory of Creativity Self-Efficacy

The Inventory of Self-Efficacy in Creativity Digital Games (Yeh & Lin, 2018) was adapted into the Inventory of Creativity Self-Efficacy (ICSE) in this study to measure the participants' level of creative self-efficacy. The adaptation only removed the situation description "during game playing." The ICSE is a 6-point Likert-type scale from 1 point to 6 points, representing strongly disagree to strongly agree. With a total of 9 items, the ICSE includes two factors: Ability to generate creative ideas (6 items) and Achievement of creative performance (3 items). Exploratory factor analysis revealed that the total variance explained by the two factors was 73.27%. The items include statements such as "I believe that I can come up with many creative ideas" and "I am more creative than most of my classmates." The Cronbach's α coefficients of the ICSE and the two factors were .927, .908, and .844, with factor loadings of .606 to .879, 73.27% of the total variance was explained by the two factors.

3.2.3. Creativity tasks

In this study, we requested the participants in the experimental groups to upload the photos they took to a designated website. For each photo, we requested the participant to write an imaginative narrative. Examples are shown as Figure 1. The imagination score, ranging from 0 points to 5 points, was rated by two trained coauthors based on their consensus. The rubrics for the scoring are as follows: 0: Roughly describes the content of the photo, but no associative thinking; 1: Associates the external features of the photo with some concrete objects/things/uses, but no descriptions of the situation; 2: Associates the concrete objects/things in the photo with self-experiences (e.g., feelings, memories, problems); 3. Associates the abstract concepts in the photo with self-experiences (e.g., feelings, memories, problems); 4: Associates the concrete objects/things or the abstract concepts in the photo with situations beyond one's own experiences; 5: Associates the concrete objects/things or the abstract concepts in the photo with situations beyond one's own experiences with vivid or touching descriptions. The rubrics were decided based on the discussion and consensus of the authors after reading 30 participants' imaginative narratives.

Observation:
I was about to eat this orange today when I noticed that it had a strange line on its side.

Imagination: What do you imagine?
This line is just like the scar from a Caesarean section left on a mother's belly. The skin sacrifices itself with no complaint; it allows the wind to blow it, the sun to scorch it, and the rain to drench it just to protect the fresh pulp inside. The wrinkles and freckles on the outside and the soft and juicy pulp inside have a strong contrast.

Figure 1. An example of uploaded photos and imaginative narratives

3.2.4. Reflection questionnaire

To understand the participants' beliefs toward the interventions of knowledge sharing and self-determination at the end of the experiment, we requested the related groups (G3 and G4) to complete a 6-point Likert-type reflection questionnaire, which included 7 items with response options from 1 point to 6 points, representing strongly disagree to strongly agree. The items are displayed in the results section.

3.3. Procedures and interventions for the experiment

Data were collected through a website designed by the researchers. For the inventory development, participants completed BMD-CL online. For the experimental instruction, our central idea was that, for the smartphone-based mindfulness intervention, which emphasized mindfulness, self-determination, and knowledge sharing online would enhance mindfulness and imagination toward surrounding things in everyday life, which would further foster creativity and creative self-efficacy. Meanwhile, BMD-CL would interact with the interventions and influence the learning process (see Figure 2 for the framework). Self-determination involves the concepts of choice, self-control, and self-management (Peterson et al., 2020). In this study, we employed the concept of self-determination by allowing participants to freely take photos

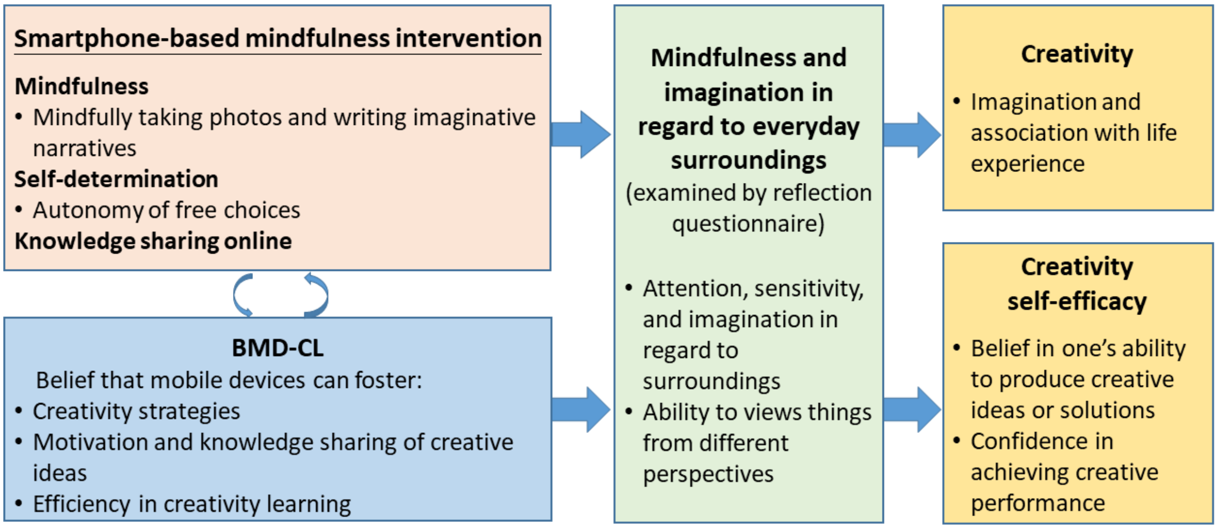


Figure 2. The framework of the experimental instruction

To further understand the influence of the intervention components, this study employed a pretest-posttest control group design that included four groups. During the one-week experimental period, the control group, Group 1, did not receive any intervention, whereas the experimental groups, Groups 2, 3, and 4, were requested to use their smartphones to take photos with different emphases for four days and to share the photos with imaginative narratives on a designated website. Group 2 had emphases on complete free choices of photo taking only, Group 3 had emphases on complete free choices of photo taking and idea sharing, and Group 4 had emphases on free choices of photo taking in varied categories and idea sharing. The restriction of categories was based on the importance of multiple-perspective thinking of mindfulness and creativity (Langer, 2016)

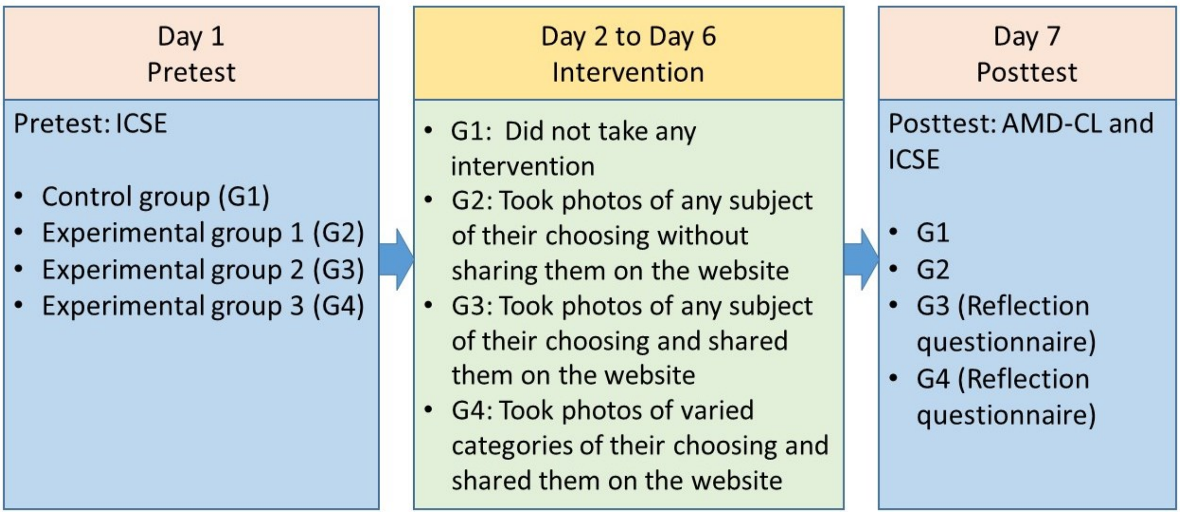


Figure 3. Procedures and interventions for the experiment

On day 1, all participants completed the pretest. From day 2 to day 6, the participants in the experimental groups took the assigned interventions; they were requested to take two photos each day, with four days in total during the 5-day period. These instructions were given under the consideration that the participants might not be able to take photos for five consecutive days. On day 7, all participants completed the posttest (see Figure 3 for detailed interventions).

4. Results

4.1. The development of BMD-CL

4.1.1. Exploratory factor analysis and reliability analysis

In this study, we used AMOS windows 21.0 to run the confirmatory factor analysis and employed SPSS windows 21.0 to run all the other analyses. Exploratory factor analysis, internal-consistency reliability, and confirmatory factor analysis were employed to examine the reliability and validity of BMD-CL. Finally, 13 items were kept in BMD-CL. Principal Component Analysis and direct oblimin were employed in factor extraction and rotation when conducting exploratory factor analysis. The results yielded three factors: strategy enhancement (5 items), motivation and knowledge sharing (4 items), and efficiency (4 items). With factor loadings ranging from .480 to .902, 69.55% of the total variance was explained by the three factors (see Table 1). The correlations between each of the factors and the total score were .919, .783, and .865 (ps < .001), respectively. The Cronbach's α of BMD-CL and the three factors were .917, .861 (strategy enhancement), .846 (Motivation and Knowledge sharing), and .870 (efficiency). Moreover, the item-total correlation coefficients ranged from .528 to .802.

4.1.2. Confirmatory factor analysis

The three-factor structure with 13 items extracted from the exploratory factor analysis was validated by confirmatory factor analysis with maximum likelihood estimation. Confirmatory factor analysis results indicated that BMD-CL has good construct validity and reliability: $\chi^2(N = 183, df = 58) = 103.395$ ($p < .05$); goodness-of-fit index = .921; root mean square residual = .046; root mean square error of approximation = .066; incremental fit index = .967; comparative fit index = .967. Moreover, the composite reliability values were .861, .851, .852, and the average variance extracted values were .557, .591, and .589 (see Table 1).

Table 1. The factor loadings of BMD-CL (N = 183)

No	Factors and items	Factor loading		
		1	2	3
Factor 1: Strategy enhancement (α = .861)				
9	Mobile devices help enhance my multi-perspective thinking.	.817		
8	Mobile devices help enhance my creative skills, such as brainstorming, storytelling, etc.	.805		
10	Mobile devices help me produce unique ideas.	.745		
13	Mobile devices help enhance my abilities of observation and sensitivity.	.735		
12	Mobile devices help me elaborate my ideas.	.573		
Factor 2: Motivation and knowledge sharing (α = .846)				
6	Mobile devices are great tools for knowledge co-creation.		.862	
3	Mobile devices help me express and share creative ideas quickly.		.861	
5	Mobile devices are great tools for collecting creative ideas.		.804	
7	Mobile devices provide many learning opportunities.		.569	
Factor 3: Efficiency (Cronbach's α = .870)				
2	Using mobile devices is an efficient way to learn creativity.			.902
1	Using mobile devices can effectively improve my creativity.			.843
4	It is easy to bring about creative ideas through mobile devices.			.677
11	Mobile devices help me produce many creative ideas quickly.			.480

4.2. Attitude toward the interventions

To understand the participants' smartphone use in everyday life, we asked them a few questions. The participants reflected that, while using smartphones during the study, they spent 1.18 hours on curriculum-related learning, 2.27 hours on personal interests unrelated to curriculum learning, 2.75 hours on social media, 2.14 hours listening to music or watching movies, and 1.06 hours playing games. These results revealed that smartphones have become an essential tool for college students' learning and social life.

This study assumed that Group 3 and Group 4 would have better learning effects for the implementation of knowledge sharing and self-determination. We designed a 6-point Likert-type reflection questionnaire to understand the participants' feelings toward the interventions. The results revealed that the participants had a positive attitude toward the interventions (see Table 2); they became more attentive, sensitive, and imaginative about things around them, suggesting an effective mindfulness intervention.

Table 2. Mean and standard deviation of the reflective questions

Question	Group 3		Group 4	
	M	*SD*	*M*	*SD*
1. This smartphone activity has made me more attentive to my surroundings.	4.82	.846	4.86	.887
2. This smartphone activity, although not lasting a long time, has enhanced my sensitivity toward my surroundings.	4.64	.859	4.59	.832
3. This smartphone activity has enhanced my ability to views things from different perspectives.	4.76	.792	4.78	.886
4. This smartphone activity has enhanced my ability to view my surroundings imaginatively.	4.70	.770	4.57	1.015
5. This smartphone activity has enhanced my everyday creativity.	4.42	.902	4.30	1.077
6. During this smartphone activity, I frequently reviewed others' uploaded assignments.	4.12	.960	4.30	.878
7. I have carefully read others' uploaded assignments.	3.79	1.023	3.95	1.104

4.3. Group differences on enhancement of creativity

To understand the learning progress of the three experimental groups, we scored each participant's creativity based on the imaginative narratives they had uploaded. Each participant had uploaded 8 photos with imaginative narratives. A Repeated Measures Analysis of Variance, with Group (G1, G2, G3, and G4) as the between variable and Day (Day 1 [D1] vs. Day 2 [D2] vs. Day 3 [D3] vs. Day 4 [D4] creativity score) as the within variable, was employed to examine group differences in creativity improvement. The results revealed a significant main effect of Day, $F(3, 103) = 26.306$, $p < .001$, $\eta^2_p = .202$; participants' creative performance on D3 and D4 was better than that on D1 and D2, and the performance on D2 was better than that on D1. In addition, there was a significant Day × Group interaction on creativity, $F(3, 103) = 5.425$, $p < .001$, $\eta^2_p = .094$. Results of simple main effect were as follows: No group differences were found on D1; on D2, G4 outperformed G2 and G3; on D3 and D4, G3 outperformed the other groups. Within each group, G2 and G3 performed better on D3 and D4 than on D1 and D2, and G4 performed better on D2–4 than on D1 (Table 3 and Figure 4).

Table 3. Group differences on enhancement of creativity

Source	ANOVA				Post hoc test
	MS	*F* (3, 103)	*p*	η^2_p	
Day	11.807	26.306***	.000	.202	D3 & D4 >D1 & D2; D2 > D1
Day x Group	2.435	5.425***	.000	.094	D2: G4 > G2, G4 > G3 D3: G3 > G4; D4: G3 > G4 G2: D3 & D4 > D1 & D2 G3: D3 & D4 > D1 & D2 G4: D2, D3 & D4 > D1
Group	2.237	2.244	.111	.041	*ns.*

Note. D1 = Day 1; D2 = Day 2; D3 = Day 3; D4 = Day 4. G1 = Group 1; G2 = Group 2; G3 = Group 3; G4 = Group 4. $^{***}p < .001$.

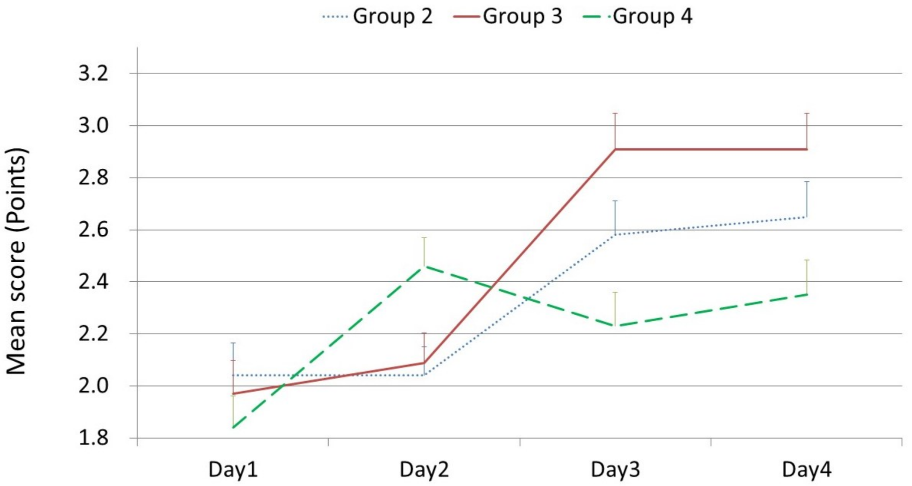


Figure 4. Mean and standard deviation of the creativity score for the three experimental groups

4.4. Effects of BMD-CL and intervention groups on creative self-efficacy

With Group A (intervention groups: G1, G2, G3, and G4) and Group B (Low and High BMD-CL) as the between variable, the pretest score of creative self-efficacy as the covariance, and the posttest score of creative self-efficacy as the dependent variable, we conducted a two-way Analysis of Covariance to examine whether beliefs toward mobile devices would influence the improvement in creative self-efficacy. The cut-points of the BMD-CL groups were the median. The results revealed a significant Group A main effect, $F(3, 129) = 2.671$, $p = .050$, $\eta^2_p = .058$, as well as a significant Group B main effect, $F(3, 129) = 6.108$, $p = .015$, $\eta^2_p = .045$. However, the Group A × Group B interaction effect was not significant. Post hoc test revealed that the experimental groups (G2, G3, and G4) had better improvement in creative self-efficacy than the control group (G1). Moreover, those who had a higher level of BMD-CL improved more in creative self-efficacy than their counterparts (Table 4 and Figure 5).

Table 4. Effects of Intervention x BMD-CL intervention on enhancement of creative self-efficacy

Source	Analysis of Covariance				Post hoc test
	MS	$F(3, 129)$	p	η^2_p	
Corrected Model	10.284	33.946***	.000	.678	
Intercept	4.360	14.393***	.000	.100	
Pretest of ICSE	66.145	218.339***	.000	.629	
Group A	.809	2.671*	.050	.058	G2, G3, G4 > G1
Group B	1.850	6.108*	.015	.045	High > Low
Group A × Group B	.144	.476	.700	.011	

Note. Group A: Intervention group; Group B: BMD-CL group. $^{*}p < .05$; $^{***}p < .001$.

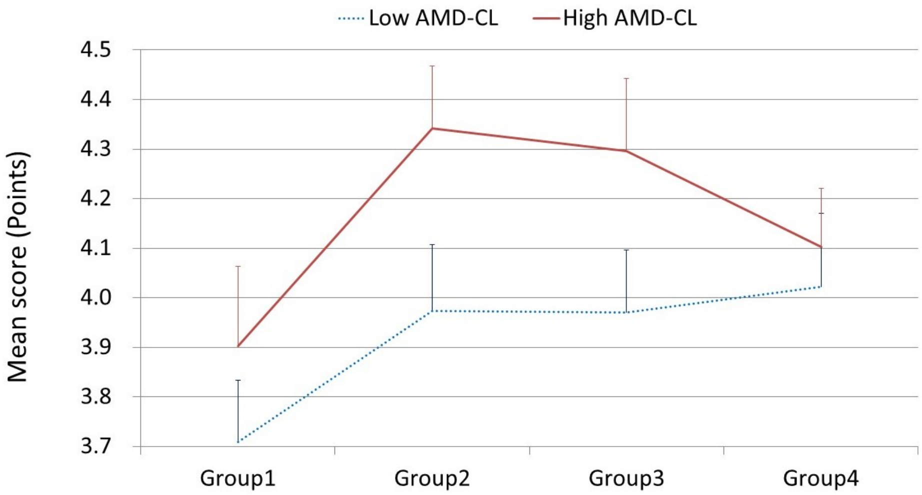


Figure 5. Mean and standard deviation of the ICSE score for BMD-CL groups in the intervention groups

5. Discussion

This study aimed at examining whether enhancing mindful learning in everyday life through varied smartphone-based interventions would improve creativity as well as whether the beliefs toward mobile-based learning would influence such a training effect. To achieve these goals, we first developed BMD-CL; then, we employed a pretest-posttest control group design to conduct an experimental instruction. The results suggest that BMD-CL has good reliability and validity. The Cronbach's α coefficient was .917, and the confirmatory factor analysis showed that the three-factor structure was a good-fit model (Goodness-fit-index = .921).

Regarding intervention effects, we employed four groups (one control group and three experimental groups with varied interventions) to test the three hypotheses we proposed. The results support H1 and suggest that practicing mindful learning in everyday life by taking photos and writing imaginative narratives can enhance college students' creativity. Specifically, we found that, overall, the participants in the experimental groups had improved their creativity during the experimental period as evidenced by the fact that their performance on D3 and D4 was better than that on D1 and D2. Moreover, they became more attentive, sensitive, and imaginative about things around them after the interventions, which was revealed through the reflection questionnaire. These results support that Langerian mindfulness has positive effects on learning (Miralles-Armenteros et al., 2019; Yeh et al., 2019a), attention (Zilcha-Mano & Langer, 2016), cognitive flexibility (Haller et al., 2017), and creativity (Grant et al., 2004; Yeh et al., 2019b). When people are in this state of mindfulness, they are actively engaged in the present, conscious of new things, and sensitive to context; moreover, mindfulness leads to benefits such as an increase in competence as well as a boost in memory, creativity, and positive affect (Langer, 2000). These cognitive mechanisms are crucial to creative performance.

To explore the optimal learning effect, this study integrated smartphone-based learning and mindful learning interventions with varied design by incorporating self-determination and/or knowledge. In addition, this study incorporated the concept of dynamic assessment by measuring the participants' performance for four days. The significant Day × Group interactions on creativity revealed that G4, which was requested to freely take photos in different categories and share photos online, outperformed the other experimental groups at the beginning, but as the interventions proceeded, G2 (emphasizing self-determination only) and G3 (emphasizing knowledge sharing and completely free choices) progressed steadily, and G3 made the best progress. These results partially support H2 and suggest that, although the autonomy of self-determination (free choices in this study) and knowledge sharing are critical to the enhancement of creativity, self-determination seems to carry more impact on such improvement than knowledge sharing. Notably, we designed G4 based on Langer's (2016) claim that mindfulness is characterized by the continuous creation of new categories, and we assumed such design would encourage multi-perspective thinking and, therefore, enhance creativity. The findings, however, suggest such design may enhance creativity for a short period of time, but in the long run, it may constrain creativity development. In addition, the fact that G2 had a better performance than G4 at the end may reflect that self-determination can better predict the improvement of creativity than knowledge sharing can. Nevertheless, the improvement of G3 suggests that when both self-determination and knowledge sharing are incorporated in smartphone-based intervention, participants have the best improvement in creativity. When integrated, the findings of this study support that self-determination helps individuals to become autonomous in their regulation and to engage in activities in a harmonious way (Deci & Ryan, 2004; Tóth-Király, Bőthe, Márki, Rigó & Orosz, 2019) and that the ideal creative process is unstructured, open-ended, and free of external limitations (Rosso, 2014). In addition, the findings of this study support that sharing ideas through observation, practice, and participation in communities contributes to professional development in creativity (Yeh, Huang, & Yeh, 2011).

Finally, the results revealed that the experimental groups (G2, G3, and G4) had better improvement in creative self-efficacy than the control group. Moreover, those who had a higher level of BMD-CL improved more in creative self-efficacy than their counterparts across groups. These findings suggest that smartphone-oriented mindfulness practiced in everyday life may improve creative self-efficacy. The findings also suggest that when people have positive feelings toward using mobile devices to improve creativity, they may be more motivated to take full advantage of mobile devices and to continue to practice enhancing their creativity. As a result of such learning, they may become mindful and enjoy the learning process, which further leads to enhancement of creativity (Boyle et al., 2011; Yeh et al., 2019a). There has been an influx of research identifying personal factors that foster creativity (Wang et al., 2018). This study, however, is the first to confirm the positive relationship between BMD-CL and creativity in smartphone-based mindfulness interventions.

6. Conclusion

Smartphones have become the most popular device for learning as well as for capturing moments in life among college students. This study pioneers the development of BMD-CL as well as the integration of smartphone use, mindfulness, knowledge sharing, self-determination, and everyday creativity to design varied types of interventions to enhance college students' ability and self-efficacy of creativity. Meanwhile, college students' beliefs toward using mobile devices to improve creativity were also considered.

The major contributions of this study are as follows. First, this study develops the "Beliefs toward Mobile Devices for Creativity Learning (BMD-CL)," a reliable and valid instrument for use in further instruction and research. Second, a creativity learning and instructional approach with ecological validity—smartphone-based mindfulness learning in everyday life—is proposed and confirmed. The findings of this study suggest that, even if only practiced for a short time, mindful learning using a smartphone lens in everyday life can effectively enhance attention, sensitivity, and imagination in regard to one's surroundings, which further leads to improvement of the ability of creativity. Moreover, practicing mindfulness in everyday life by taking photos and writing imaginative narratives, especially when both self-determination and knowledge sharing are emphasized, can enhance college students' creativity. Finally, the findings suggest that beliefs toward using mobile devices to enhance creative self-efficacy play an important role, which reminds researchers of the importance of a positive belief toward mobile devices' potential for creativity learning.

Notably, since this study was conducted in an everyday life situation, it has better ecological validity than those conducted in laboratories. This study contributes to providing a very convenient and feasible approach for enhancing personal creativity through smartphones and computers, which provides insights for the instructional design of creativity learning.

7. Limitations, Implications, and further studies

The results showed that the participants spent about 3.45 hours learning and 2.75 hours on social media using their smartphones per day, suggesting that smartphones have become an essential tool for college students' learning and social life. Implementing mindfulness interventions through mobile devices to enhance creativity is a promising approach for mindfulness, and creativity is highly valued for school education and life-long learning. This study included four groups to compare the effectiveness of varied interventions. Because of the difficulty of finding enough volunteers to engage in the experiment for a longer period of time, our interventions only lasted a week. Further studies may include our intervention design as part of the course requirement and implement it for a longer period of time to examine its effects and to maximize the intervention effect.

Most of the recent studies involving self-determination and smartphones have focused on the problems brought about by the lack of control in self-determination (e.g., Gugliandolo, Costa, Kuss, Cuzzocrea, & Verrastro, 2019; Long et al., 2019). Few studies, however, focus on how to boost creativity while taking advantage of self-determination and smartphones. This study pioneers the implementation of such intervention research in everyday life. Future studies can continue to explore different intervention effects. Such studies should provide valuable information for educators who attempt to improve students' creativity through mobile devices.

Finally, this study found that beliefs toward using mobile devices are influential to the enhancement of creativity self-efficacy in our experiment. Future studies can include a bigger sample to confirm such a relationship through investigation studies, or to include other personal traits to examine their influences on interventions using smartphones. Future research can also try to replicate the experimental design using other mobile devices.

Acknowledgment

This work was supported by the Ministry of Science and Technology, Taiwan (Grant No. MOST 107-2511-H-004-002-MY2).

References

Ashford, R., Giorgi, S., Mann, B., Sherritt, L., Ungar, L., & Curtis, B. (2020). Digital recovery networks: Characterizing user participation, engagement, and outcomes of a novel recovery social network smartphone application. *Journal of Substance Abuse Treatment, 109*, 50-55. doi:10.1016/j.jsat.2019.11.005

Autin, F., & Croizet, J. (2012). Improving working memory efficiency by reframing metacognitive interpretation of task difficulty. *Journal of Experimental Psychology: General, 141*(4), 610-618. doi:10.1037/a0027478

Augello, A., Infantino, I., Lieto, A., Pilato, G., Rizzo, R., & Vella, F. (2016). Artwork creation by a cognitive architecture integrating computational creativity and dual process approaches. *Biologically Inspired Cognitive Architectures*, *15*, 74-86. doi:10.1016/j.bica.2015.09.007

Baas, M., Nevicka, B., & Velden, F. S. T. (2020). When paying attention pays off: The Mindfulness skill act with awareness promotes creative idea generation in groups. *European Journal of Work and Organizational Psychology, 29*(4), 619-632. doi:10.1080/1359432X.2020.1727889

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

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

Bandura, A. (2001). Social cognitive theory: An Agentic perspective. *Annual Review of Psychology*, *52*(1), 1-26. doi:10.1146/annurev.psych.52.1.1

Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Journal of Management, 38*, 9-44. doi:10.1177/0149206311410606

Ben-Ari, R. (2002). Mindful movement: Towards enhanced intergroup relations in heterogeneous classrooms. *Curriculum and Teaching, 17*(2), 19-36. doi:10.1093/elt/56.2.128

Bercovitz, K., Pagnini, F., Phillips, D., & Langer, E. (2017). Utilizing a creative task to assess Langerian mindfulness. *Creativity Research Journal*, *29*(2), 194-199. doi:10.1080/10400419.2017.1304080

Bishop, S. R., Lau, M., Shapiro, S., Carlson, L., Anderson, N. D., Carmody, J., Segal, Z. V., Abbey, S., Speca, M., Velting, D., & Devins, G. (2004). Mindfulness: A Proposed operational definition. *Clinical Psychology: Science and Practice, 11*, 230-241. doi:10.1093/clipsy.bph077

Boyle, E., Connolly, T. M., & Hainey, T. (2011). The Role of psychology in understanding the impact of computer games. *Entertainment Computing, 2*(2), 69-74. doi:10.1016/j.entcom.2010.12.002

Caldwell, K., Harrison, M., Adams, M., Quin, R. H., & Greeson, J. (2010). Developing mindfulness in college students through movement based courses: Effects on self-regulatory self-efficacy, mood, stress, and sleep quality. *Journal of American College Health, 58*(5), 433-442. doi:10.1080/07448480903540481

Creswell, J. D. (2017). Mindfulness interventions. *Annual Review of Psychology*, *68*, 491-516. doi:10.1146/annurev-psych-042716-051139

Davenport, C., & Pagnini, F. (2016). Mindful learning: A Case study of Langerian mindfulness in schools. *Frontier in Psychology, 7*, 1372. doi:10.3389/fpsyg.2016.01372

Deci, E. L., & Ryan, R. M. (Eds.). (2004). *Handbook of self-determination research*. Rochester, NY: University of Rochester Press.

Farley, H., Murphy, A., Johnson, C., Carter, B., Lane, M., Midgley, W., Dekeyser, S., & Koronios, A. (2015). How do students use their mobile devices to support learning? A Case study from an Australian regional university. *Journal of Interactive Media in Education, 14*(1), 1-13. doi:10.5334/jime.ar

Gärtner, C. (2013). Enhancing readiness for change by enhancing mindfulness. *Journal of Change Management, 13*(1), 52-68. doi:10.1080/14697017.2013.768433

Gong, Y., Huang, J. C., & Farh, J. L. (2009). Employee learning orientation, transformational leadership, and employee creativity: The mediating role of employee creative self-efficacy. *Academy of management Journal*, *52*(4), 765-778. doi:10.5465/amj.2009.43670890

Grant, A. M., Langer, E. J., Falk, E., & Capodilupo, C. (2004). Mindful creativity: Drawing to draw distinctions. *Creativity Research Journal, 16*(2-3), 261-265. doi:10.1080/10400419.2004.9651457

Greason, P. B., & Cashwell, C. S. (2009). Mindfulness and counseling self-efficacy: The Mediating role of attention and empathy. *Counselor, 49*(1) 2-19. doi:10.1002/j.1556-6978.2009.tb00083.x

Gu, C., Hu, B. Y., Ngwira, F. F., Jing, Z., & Zhou, Z. (2016). The Effect of general creative personality and freedom of task choice on adolescents' social creativity. *The Journal of Creative Behavior*, *50*(2), 132-149. doi:10.1002/jocb.74

Gugliandolo, M. C., Costa, S., Kuss, D. J., Cuzzocrea, F., & Verrastro, V. (2019). Technological addiction in adolescents: The interplay between parenting and psychological basic needs. *International Journal of Mental Health and Addiction*, 1-14. doi:10.1007/s11469-019-00156-4

Haller, C. S., Bosman, C. M., Kapur, K., Zafonte, R., & Langer, E. J. (2017). Mindful creativity matters: Trajectories of reported functioning after severe traumatic brain injury as a function of mindful creativity in patients' relatives: A Multilevel analysis. *Quality of life research*, *26*(4), 893-902. doi:10.1007/s11136-016-1416-1

Hu, J., Erdogan, B., Jiang, K., Bauer, T. N., & Liu, S. (2018). Leader humility and team creativity: The Role of team information sharing, psychological safety, and power distance. *Journal of Applied Psychology*, *103*(3), 313-323. doi:10.1037/apl0000277

Iani, L., Lauriola, M., Cafaro, V., & Didonna, F. (2017). Dimensions of mindfulness and their relations with psychological well-being and neuroticism. *Mindfulness*, *8*(3), 664-676. doi:10.1007/s12671-016-0645-2

Kabat-Zinn, J. (1994). *Wherever you go, there you are: Mindfulness meditation in everyday life*. New York, NY: Hyperion Books.

Kremer, H., Villamor, I., & Aguinis, H. (2019). Innovation leadership: Best-practice recommendations for promoting employee creativity, voice, and knowledge sharing. *Business Horizons*, *62*(1), 65-74. doi:10.1016/j.bushor.2018.08.010

Langer, E. J. (1989). *Mindfulness*. Cambridge, MA: De Capo Press.

Langer, E. J. (2000). Mindful learning. *Current Directions in Psychological Science, 9*(6), 220-223. doi:10.1111/1467-8721.00099

Langer, E. (2012). The mindless use of medical data. *Journal of Business Research*, *65*(11), 1651-1653. doi:10.1016/j.jbusres.2011.02.018

Langer, E. J. (2016). *The Power of mindful learning*. Boston, MA: Da Capo Lifelong Books.

Lawrie, E. J., Tuckey, M. R., & Dollard, M. F. (2018). Job design for mindful work: The Boosting effect of psychosocial safety climate. *Journal of Occupational Health Psychology*, *23*(4), 483-495. doi:10.1037/ocp0000102

Levy, B. R., Jennings, P., & Langer, E. J. (2001). Improving attention in old age. *Journal of Adult Development*, *8*(3), 189-192. doi:10.1023/A:1009546408665.

Long, J., Wang, P., Liu, S., & Lei, L. (2019). Materialism and adolescent problematic smartphone use: The Mediating role of fear of missing out and the moderating role of narcissism. *Current Psychology*. doi:10.1007/s12144-019-00526-0

Men, C., Fong, P. S., Luo, J., Zhong, J., & Huo, W. (2019). When and how knowledge sharing benefits team creativity: The Importance of cognitive team diversity. *Journal of Management & Organization*, *25*(6), 807-824. doi:10.1017/jmo.2017.47

Miralles-Armenteros, S., Chiva-Gómez, R., Rodríguez-Sánchez, A., & Barghouti, Z. (2019). Mindfulness and academic performance: The Role of compassion and engagement. *Innovations in Education and Teaching International*, 1-11. doi:10.1080/14703297.2019.1676284

Ostafin, B. D., & Kassman, K. T. (2012). Stepping out of history: Mindfulness improves insight problem solving. *Consciousness and Cognition, 21*(2), 1031-1036. doi:10.1016/j.concog.2012.02.014

Pang, D., & Ruch, W. (2019). The mutual support model of mindfulness and character strengths. *Mindfulness, 10*, 1545-1559. doi:10.1007/s12671-019-01103-z

Peterson, M., Aljadeff-Abergel, E., Eldridge, R. R., VanderWeele, N. J., & Acker, N. S. (2020). Conceptualizing self-determination from a behavioral perspective: The Role of choice, self-control, and self-management. *Journal of Behavioral Education*. doi:10.1007/s10864-020-09368-4

Polsinelli, A., J., Kaszniak, A. W., Glisky, E., & Ashish, D. (2020). Effects of a brief, online, focused attention mindfulness training on cognition in older adults: A Randomized controlled trial. *Mindfulness, 11*, 1182-1193. doi:10.1007/s12671-020-01329-2

Rosso, B. D. (2014). Creativity and constraints: Exploring the role of constraints in the creative processes of research and development teams. *Organization Studies, 35,* 551-585. doi:10.1177/0170840613517600

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67. doi:10.1006/ceps.1999.1020

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness.* New York, NY: Guilford Press.

Stewart, C., & Bower, M. (2019). Novice online educator conceptual frameworks: A Mental model exploration of mindful learning design. *Educational Media International*, *56*(1), 14-43. doi:10.1080/09523987.2019.1583463

Sung, S. Y., & Choi, J. N. (2020). Effects of diversity on knowledge sharing and creativity of work teams: Status differential among members as a facilitator. *Human Performance, 32*(3-4), *145-164.* doi:10.1080/08959285.2019.1639712

Tarrasch, R., Berger, R., & Grossman, D. (2020). Mindfulness and compassion as key factors in improving teacher's well-being. *Mindfulness, 11*, 1049–1061. doi:10.1007/s12671-020-01304-x

Tierney, P., & Farmer, S. M. (2002). Creative self-efficacy: Its potential antecedents and relationship to creative performance. *Academy of Management Journal, 45*(6), 1137-1148. doi:10.2307/3069429

Tóth-Király, I., Bőthe, B., Márki, A. N., Rigó, A., & Orosz, G. (2019). Two sides of the same coin: The Differentiating role of need satisfaction and frustration in passion for screen-based activities. *European Journal of Social Psychology, 49*(6), 1190-1205. doi:10.1002/ejsp.2588

Vansteenkiste, M., & Ryan, R. M. (2013). On psychological growth and vulnerability: Basic psychological need satisfaction and need frustration as a unifying principle. *Journal of Psychotherapy Integration, 23*(3)*,* 263-280. doi:10.1037/a0032359

Vartanian, O. (2019). Fluid intelligence, working memory, and creativity: A Componential view. In S. Nalbantian & P. M. Matthews (Eds.), *Secrets of creativity: What neuroscience, the arts, and our minds reveal* (pp. 45–63). New York, NY: Oxford University Press.

Wang, S., Liu, Y., & Shalley, C. E. (2018). Idiosyncratic deals and employee creativity: The Mediating role of creative self-efficacy. *Human Resource Management*, *57*(6), 1443-1453. doi:10.1002/hrm.21917

Wang, Y., & Liu, C. (2016). Cultivate mindfulness: A Case study of mindful learning in an English as a foreign language classroom. *IAFOR Journal of Education*, *4*(2), 141-155. doi:10.22492/ije.4.2.08

Woolfolk, A. (2016). *Educational psychology* (13th ed.). Boston, MA: Allyn & Bacon.

Yeh, Y. C., & Lin, C. S. (2018). Achievement goals influence mastery experience via two paths in digital creativity games among elementary school students. *Journal of Computer Assisted Learning*, *34*(3), 223-232. doi:10.1111/jcal.12234

Yeh, Y. C., Chang, H. L., & Chen, S. Y. (2019a). Mindful learning: A Mediator of mastery experience during digital creativity game-based learning among elementary school students. *Computers & Education*, *132*, 63-75. doi:10.1016/j.compedu.2019.01.001

Yeh, Y. C., Chen, S. Y. C., Rega, E. M. R., & Lin, C. S. L. (2019b). Mindful learning experience facilitates mastery experience through heightened flow and self-efficacy in game-based creativity learning. *Frontiers in Psychology*, *10*, 1593. doi:10.3389/fpsyg.2019.01593

Yeh, Y. C., Huang, L. Y., & Yeh, Y. L. (2011). Knowledge management in blended learning: Effects on professional development in creativity instruction. *Computers & Education, 56*(1), 146-156. doi:10.1016/j.compedu.2010.08.011

Yeh, Y. C., Yeh, Y. L., & Chen, Y. H. (2012). From knowledge sharing to knowledge creation: A Blended knowledge-management model for improving university students' creativity. *Thinking Skills and Creativity*, *7*(3), 245-247. doi:10.1016/j.tsc.2012.05.004

Yeh, Y. C., Tsai, J. L., Hsu, W. C., & Lin, C. F. (2014). A model of how working memory capacity influences insight problem solving in situations with multiple visual representations: An eye tracking analysis. *Thinking Skills and Creativity*, *13*, 153-167. doi:10.1016/j.tsc.2014.04.003

Yeh, Y., & Lin, C. F. (2015). Aptitude-treatment interactions during creativity training in e-learning: How meaning-making, self-regulation, and knowledge management influence creativity. *Educational Technology & Society, 18*(1), 119-131.

Yeh, Y., Lai, G. J., Lin, C. F., Lin, C. W., & Sun, H. C. (2015). How stress influences creativity in game-based situations: Analysis of stress hormone, negative emotions, and working memory. *Computers & Education, 81*, 143-153. doi:10.1016/j.compedu.2014.09.011

Yeh, Y., Lai, S. C., & Lin, C. W. (2016). The dynamic influence of emotions on game-based creativity: An integrated analysis of emotional valence, activation strength, and regulation focus. *Computers in Human Behavior, 55*, 817-825. doi:10.1016/j.chb.2015.10.037

Zhang, W., Sun, S. L., Jiang, Y., & Zhang, W. (2019). Openness to experience and team creativity: Effects of knowledge sharing and transformational leadership. *Creativity Research Journal, 31*(1), 62–73. doi:10.1080/10400419.2019.1577649

Zheng, X., & Liu, X. (2017). The Buffering effect of mindfulness on abusive supervision and creative performance: A social cognitive framework. *Frontiers in Psychology*, *8*, 1588. doi:10.3389/fpsyg.2017.01588

Zilcha-Mano, S., & Langer, E. (2016). Mindful attention to variability intervention and successful pregnancy outcomes. *Journal of Clinical Psychology*, *72*(9), 897-907. doi:10.1002/jclp.22294

Zimmerman, B. J. (2011). Motivational sources and outcomes of self-regulated learning and performance. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 49-64). New York, NY: Routledge.

Shadiev, R., Zhang Z. H., Wu, T.-T., & Huang, Y. M. (2020). Review of Studies on Recognition Technologies and Their Applications Used to Assist Learning and Instruction. *Educational Technology & Society*, *23* (4), 59–74.

Review of Studies on Recognition Technologies and Their Applications Used to Assist Learning and Instruction

Rustam Shadiev[1], Zi Heng Zhang[1], Ting-Ting Wu[2] and Yueh Min Huang[3*]
[1]School of Education Science, Nanjing Normal University, China // [2]National Yunlin University of Science and Technology, Taiwan // [3]Department of Engineering Science, National Cheng Kung University, Taiwan //
rustamsh@gmail.com // 1040303326@qq.com // ttwu@yuntech.edu.tw // huangncku@gmail.com
[*]Corresponding author

(Submitted February 17, 2020; Revised April 6, 2020; Accepted September 9, 2020)

ABSTRACT: We reviewed studies on recognition technologies published in the last ten years. This review study was aimed toward identifying, appraising, selecting, and synthesizing all high quality research evidence published in the literature related to recognition technologies and on determining how they can assist learning and instruction. This study particularly focuses on summarizing the current state of knowledge in the following dimensions: (1) recognition technology and processes, (2) applications, (3) schemes, and (4) advantages and disadvantages. We reviewed seventy-two papers and identified eighteen recognition technologies. Our results showed that all of the recognition technologies under consideration featured different recognition processes and applications. In most studies, the participants were university students. Questionnaires and tests were the most frequently used data collection methods. Most studies used a group comparison as their research design. Finally, several advantages and disadvantages of the recognition technologies were identified and summarized in the papers. The most frequently cited disadvantage was a low recognition accuracy rate. Based on our results, several suggestions and implications are made for the teaching and research community.

Keywords: Review, Recognition technology, Learning and instruction

1. Introduction

Signal processing has recently become an increasingly advancing field due to rapid development of information and communication technologies. Signal processing is a branch of electrical engineering defined as a process in which information is extracted from a signal (Shih, 2010). According to van De Ville, Demesmaeker, and Preti (2017), signal processing focuses on analyzing, modifying, and synthesizing signals such as sounds, images, and biological measurements. Signal processing manipulates information content in signals and is intended to facilitate automatic recognition.

Recognition technologies are vital signal processing applications (Dong, Wang, & Kuang, 2014; Shadiev, Hwang, Chen & Huang, 2014). These technologies extract information from signals, analyze it, and then translate it into recognizable output. One example of such a technology is speech-to-text recognition. This technology receives voice input from a speaker, processes it, and then generates textual output (Shadiev et al., 2014). It has been suggested that recognition technologies can be useful with assisting learning and instruction. In the field of education, there are different types of input signals, e.g., speech, facial images, gesture images/videos, or physiological signals. The purpose of these signals is to collect data from learners and translate the data into recognizable learning data either in cognitive or affective domains. There are currently many recognition technologies, e.g., emotion recognition (Lin, Su, Chao, Hsieh, & Tsai, 2016), hand movement recognition (Wolski & Jagodziński, 2019), physical activity recognition (Lindberg, Seo, & Laine, 2016), gesture recognition (Yang & Liao, 2014), handwriting recognition (Yanikoglu, Gogus, & Inal, 2017), facial recognition (Yueh, Lin, Liu, Shoji, & Minoh, 2014), haptic recognition (Magana & Balachandran, 2017), touch recognition (Mercier, Vourloumi, & Higgins, 2017), image recognition (Hwang, Chang, Chen, & Chen, 2018), sign recognition (Ditcharoen, Naruedomkul, & Cercone, 2010), attention recognition (Chen, Wang, & Yu, 2017), biometric information recognition (Kaklauskas et al., 2010), math expression recognition (Pacheco-Venegas, López, & Andrade-Aréchiga, 2015), body position and movement recognition (Wang, Hwang, Li, Chen, & Manabe, 2019), etc. Scholars have applied such technologies to aid both learning and instruction.

There are currently many review studies on different recognition technologies. Calvo and Nummenmaa (2016) reviewed studies on facial expression recognition to explore the role of visual and emotional factors in the recognition process. The results showed that these factors contribute differently to the process; however, facial expression recognition depends on perceptual factors more than affective factors. Ghosh, Dube, and Shivaprasad (2010) synthesized results from script recognition studies focusing on different script recognition process methodologies. Their results revealed two distinct script recognition techniques: structure-based and visual appearance-based, and two different script features for script identification: character-wise and word-wise. Khalil, Jones, Babar, Jan, Zafar, and Alhussain (2019) reviewed studies on speech emotion recognition based on deep learning techniques. Khalil et al. (2019) identified deep learning techniques, such as the deep Boltzmann machine, recurrent neural networks, recursive neural networks, deep belief networks, convolutional neural networks, and auto encoders, and suggested that they offer easy model training and the efficiency in terms of shared weights. Liu and Wang (2018) focused on studies on gesture recognition in their review research. Liu and Wang (2018) covered gesture recognition technologies and algorithms and discussed how gesture recognition can be applied to human-robot collaborative manufacturing. Recognition technologies have great potential to assist learning and instruction. However, based on an analysis of the review studies referenced above, it can be concluded that they focused on one specific recognition technology only, and none of them considered recognition technologies in an educational context. To address this gap in the field, the present review study was carried out. In this review study, we identified different recognition technologies, described their recognition process, and explained their applications to learning and instruction with corresponding examples. Furthermore, schemes of the reviewed studies (e.g., participants, data collection, or research design) were also covered and the advantages and disadvantages of recognition technologies were discussed. This information may be useful for educators and researchers, especially those with a non-technical background, in terms of providing a greater understanding of recognition technologies and their recognition processes, applications, and schemes, as well as their advantages and disadvantages. Furthermore, this review study may provide ideas for future research on and development of the use of recognition technologies in the field of education.

2. Methodology

We conducted a systematic review in this study (Grant & Booth, 2009). The methodology was based on general recommendations proposed in related review studies (Shadiev & Sintawati, 2020a; Shadiev & Yang, 2020c). The methodology included three steps: a search, expert meetings, and a content analysis (Shadiev et al., 2014). First, we searched related studies on recognition technologies and their applications intended to assist learning and instruction (Figure 1). We searched such studies using the Web of Science database because Caseiro and Santos (2018) suggested that "it is one of the most extensive, popular and relevant research databases for the academic community" (p. 8). Our search was carried out in January 2020 and included such keywords as recognition, technology, learning, teaching, and instruction in different combinations because they were frequently used in earlier review studies (Shadiev et al., 2014). A total of 2,465 articles were returned by the search. After that, we used the following criteria for the purpose of narrowing down the selection of the research articles: studies published (1) within a last decade, between 2009 and 2019, (2) those focused on recognition technology applied to learning and instruction, (3) studies in English, and (4) studies appearing as a full text article in Social Science Citation Index (SSCI) journals related to educational technology. Duman, Orhon and Gedik (2015) argued that "SSCI journals adopt stringent criteria in reviewing articles. These articles are generally regarded as having higher impacts in the field" (p. 200). A total of seventy-two papers were selected and reviewed, and eighteen different recognition technologies were identified (see Appendix I).

Two researchers were involved in article selection process. First, they independently examined all articles against the above-mentioned criteria and removed all duplicate articles. After that, differences in the screening processes were identified and discussed. Finally, a consensus was reached among the researchers. We measured interrater reliability, which was high, with correlations of .95 between the researchers.

Second, the expert meeting was arranged following the general recommendations of Voogt and Roblin (2010). Eight experts were invited to the meeting. Each expert had a PhD degree and research expertise in educational technologies. The experts were asked to review eighteen different recognition technologies (see Appendix I) in order to discuss their theoretical and practical value as well as their features and functions. Based on the expert meeting, the following dimensions were derived: (1) recognition process, (2) application, (3) scheme, and (4) advantages and disadvantages.

Finally, a content analysis of seventy-two papers was carried out. It was based on open coding, and two researchers were involved in the process (Shadiev et al., 2014; Shadiev & Sintawati, 2020a; Shadiev & Yang, 2020c). The researchers carried out the coding process independently. They read the article and coded the data of interest with respect to the derived dimensions, including the recognition process, application of the recognition technology, scheme, and advantages and disadvantages of the given recognition technology. After that, the researchers finalized the coding phase through resolving any disagreements by re-examining the papers.

In the following section, we report our results and describe them with examples from the representative papers in order to enhance the reader's comprehension.

Main source of articles: *Web of Science database*

Key words: *recognition, technology, learning, teaching, and instruction*

Screening criteria:
Articles published between 2009 and 2019
Articles focused on recognition technology
Articles published in English
Articles published in journals indexed by Social Science Citation Index (SSCI)

SSCI Journals (72 studies being reviewed)
British Journal of Educational Technology (n=12)
Computers & Education (n=8)
Interactive Learning Environments (n=8)
Journal of Computer Assisted Learning (n=8)
Educational Technology & Society (n=6)
IEEE Transactions on Learning Technologies (n=6)
Computer Assisted Language Learning (n=5)
Educational Technology Research & Development (n=4)
International Journal of Computer-Supported Collaborative Learning (n=3)
Reading & Writing Quarterly (n=2)
Australasian Journal of Educational Technology (n=1)
Foreign Language Annals (n=1)
International Journal of Science and Mathematics Education (n=1)
Journal of Educational Computing Research (n=1)
Journal of Research in Science Teaching (n=1)
Learning and Instruction (n=1)
Library & Information Science Research (n=1)
ReCALL (n=1)
System (n=1)
The Internet and Higher Education (n=1)

Figure 1. Search of articles

3. Results

3.1. Recognition technology and process

Figure 2 demonstrates the number of articles published by year. The highest number of articles was published in the last four years (i.e., in 2016 (n = 11), 2017 (n = 11), 2018 (n = 13) and 2019 (n = 10)), whereas the lowest number was published in 2009 (n = 1). The publication trendline demonstrates that the trend in publishing articles is increasing.

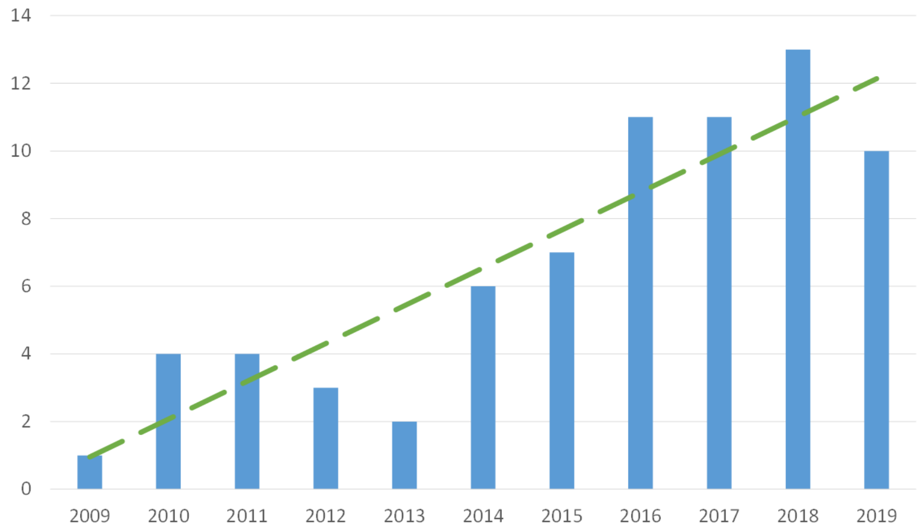


Figure 2. Published articles and publication trendline (broken green line)

Eighteen different recognition technologies were identified (see Figure 3). The most frequently used recognition technologies were speech recognition (n = 16) and touch recognition (n = 16) whereas the least frequently used were attention recognition (n = 1), biometric information recognition (n = 1), face recognition (n = 1), hand movement recognition (n = 1), math expression recognition (n = 1), physical activity recognition (n = 1), sign recognition (n = 1) and text-to-speech recognition (n = 1).

All identified recognition technologies could be grouped into six categories based on the type of input data in the recognition processes: audio, non-text image, text image, moving image, physiological signals, and other (see Appendix I). In the following, we list these technologies and explain them in relation to the recognition processes.

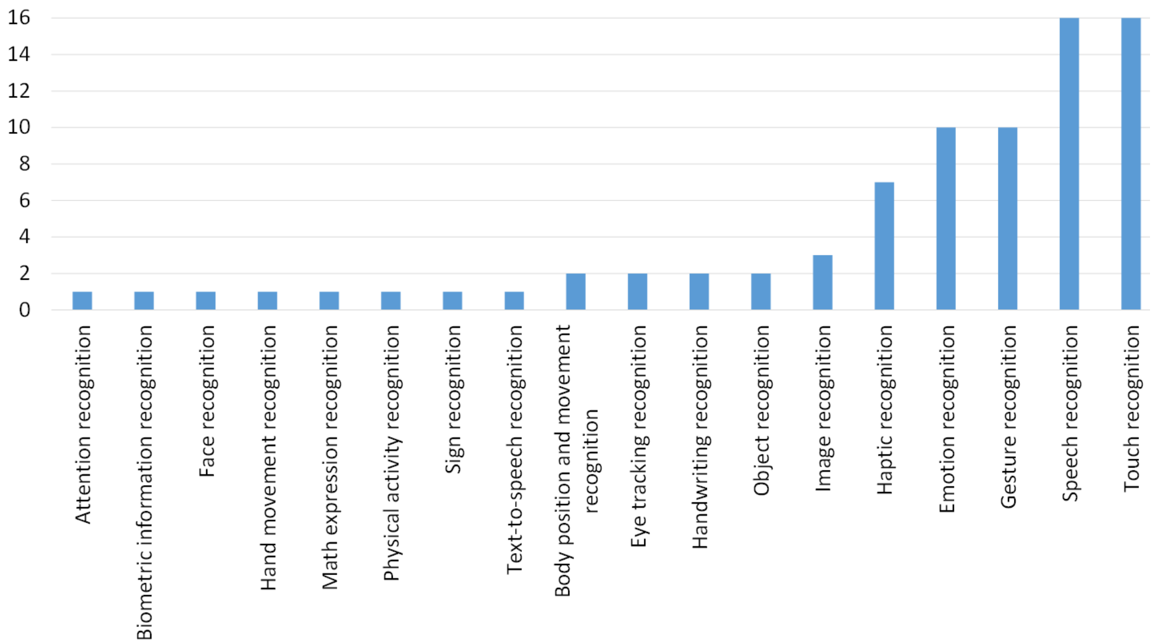


Figure 3. Recognition technologies

Audio. This category includes three types of recognition technologies based on audio: biometric information, speech, and emotion recognition technologies. Biometric information recognition technology records a student's voice using a microphone and then infers deception from stress based on a biometric voice stress analysis, i.e., the non-verbal, low-frequency content of the voice conveys information about the physiological and psychological state of the student (Kaklauskas et al., 2010). Similarly, emotion recognition technology receives voice input through a microphone and identifies the emotional states of students based on voice intonation (Bahreini, Nadolski & Westera, 2016). Speech-to-text recognition technology synchronously transcribes text streams from speech input that are simultaneously shown to students on a whiteboard or computer screens during lectures (Shadiev et al., 2014).

Non-Text image. Recognition technologies in this category are based on non-text images and include five types: face, image, sign, object, and emotion recognition technologies. For example, facial recognition technology captures the human face using a camera and compares the captured image with data stored in a database (Yueh et al., 2014). Image recognition technology detects marker labels or graphics through a camera, and then, relevant virtual information is displayed on the device when a target is recognized (Hwang et al., 2018). Sign recognition technology converts user input sign language into language text (Ditcharoen, Naruedomkul, & Cercone, 2010).

Text image. In this category, recognition technologies are based on text images and include handwriting and math expression recognition technologies. Handwriting recognition technology supports handwritten input through tablets, detects the results of student writing, and provides feedback (Yanikoglu et al., 2017). Math expression recognition technology recognizes written math symbols, converts them into formats that a computer can understand, and then displays the recognized math expressions on a computer screen (Pacheco-Venegas et al., 2015).

Moving image. This category consists of four recognition technologies that are based on moving images, i.e., body position and movement, gesture, eye tracking, and hand movement recognition technologies. Body position and movement recognition technology detects postures or movements made by students using Kinect, determines whether these postures or movements meet the requirements, and provides feedback (Wang et al., 2019). Gesture recognition technology captures live images of a user through a camera and recognizes user gestures from the live images through computer vision technology intended to observe gesture interactions (Yang & Liao, 2014). Hand movement recognition technology records and processes a student's hand movement and responses to the movement through a Kinect sensor (Wolski & Jagodziński, 2019). Eye-tracking technology measures eye positions and eye movement to reveal where visual sensory information is being acquired and how much attention a student is paying during the learning process (Huang et al., 2015; Wiebe, Minogue, Jones, Cowley & Krebs, 2009).

Physiological signals. Three types of recognition technologies based on physiological signals are included in this category: attention, emotion, and physical activity. Attention recognition technology monitors the brainwaves of a student's mental activity, registers electrical impulses generated by the brain, and outputs a value representing the current state of attention (Chen et al., 2017). Emotion recognition technology senses a student's pulse signals through an ear sensor, analyzes variability in heart rate patterns, interprets the data, and notifies a student about his/her status and changes in emotional states (Chen & Wang, 2011). Physical activity recognition technology detects a student's physical activity through the sensors in a smart phone and a smart wrist band. This technology collects data such as heart rate through the optical heart rate sensor in the smart wrist band, and then this data is recorded into a file (Lindberg et al., 2016).

Others. This category includes four recognition technologies: haptic, touch, emotion, and text-to-speech. Haptic recognition technology enables students to manipulate virtual objects through a haptic interface and then causes students to feel the force feedback (Magana & Balachandran, 2017). Touch recognition technology allows users to directly touch and interact with content without having to use an input device (Mercier et al., 2017). The technology senses the touch of a user through a camera and executes a respective command or sends respective feedback.

3.2. Application

The results show that recognition technologies were applied to educational contexts in three different ways: (1) as a tool or method to assist learning, (2) to analyze learning behavior, and (3) to identify the learning status of students (see Appendix II).

A tool or method to assist learning. In this dimension, recognition technologies are reviewed in terms of their applications as tools or methods to assist learning. Researchers have studied the application of speech recognition to speaking and pronunciation practice in foreign or second languages. For example, scholars developed an application with automatic speech recognition to improve language learners' speaking skills (Ahn & Lee, 2016; Chen et al., 2016; McCrocklin, 2016; Mroz, 2018; Van Doremalen et al., 2016), where the application provides immediate feedback regarding how correct the language learners' pronunciation is. Furthermore, language learners can obtain remedial exercises based on the type of mistakes they make.

Scholars have used visuo-haptic simulators to teach different subjects by combining tactile and visual perception, e.g., mechanics concepts in Neri et al. (2018) and Wiebe et al. (2009), physics concepts in Hamza-Lup et al. (2010), a dental curriculum in San Diego et al. (2012), or robot programming concepts in Howard, Park and Remy (2012). Learners were able to change the relevant parameters as well as apply and feel forces in the experiments. They exerted forces by manipulating a haptic stylus. Scholars have argued that haptic feedback is helpful to students to conceptualize abstract, difficult concepts (Magana & Balachandran, 2017).

Many researchers have studied touch recognition technology. This technology has been applied to assist learning of art (Hung et al., 2015), math (Chen et al., 2016; Hwang, Shadiev, Tseng & Huang, 2015; Mercier, Higgins & Da Costa, 2014), geography (Hung et al., 2014), and science (Davis et al., 2015; Horn et al., 2016). In most studies, a multi-touch tabletop technology was employed to facilitate collaboration among learners (Ardito et al., 2013; Bause et al., 2018; Mercier et al., 2014; Schmitt & Weinberger, 2019). Some scholars compared the learning effects of touch recognition technology with other approaches, e.g., pen and paper (Higgins, Mercier, Burd & Joyce-Gibbons, 2012; Mercier et al., 2017) or tablet PCs (Hwang et al., 2015). They found that a multi-touch tabletop had a greater learning effect than other approaches (Hwang et al., 2015). For example, students had better interactions and idea development under multi-touch conditions (Mercier et al., 2017).

Image recognition technology also received considerable attention in reviewed studies (Hwang et al., 2018; Liou, Yang, Chen & Tarng, 2017; Shadiev, Wu & Huang, 2020b). For example, a mobile learning system with an image recognition function was developed in Hwang et al. (2018) and Liou et al. (2017). It used image-based AR technology to support the learning of a local culture (Hwang et al., 2018) or science (Liou et al., 2017). A learner had to direct the system to a learning target, and when it was recognized, the information related to the target was displayed on a mobile device to guide learners or provide additional learning content.

Handwriting recognition technology was applied to handwriting diagnoses and remedial instruction in Hsiao et al. (2015) and Yanikoglu et al. (2017). The technology could detect errors in handwriting and provide immediate feedback. A biometric, intelligent self-assessment system based on a voice stress analysis and a special algorithm was used in Kaklauskas et al. (2010) to assist students with self-assessment. Face recognition technology was used in Yueh et al. (2014) to develop a learner identification system in order to support real-time interaction in an international distance course. The facial recognition component of the system was used to recognize learners. Wolski and Jagodzinsk (2019) developed a virtual chemical laboratory based on hand movement recognition technology. It detected the hand movement of students through Kinect, where the students could simulate chemical experiment activities in a virtual laboratory, such as operating specific laboratory equipment.

Pacheco-Venegas et al. (2015) used an automatic evaluation system for mathematical expressions. The system could recognize mathematical symbols written by students, evaluate the math expression, and finally give feedback. Ditcharoen et al. (2010) developed a system that was able to translate from Thai sign language into Thai text to help deaf students learn the Thai language. van Laere et al. (2017) used bilingual text-to-speech technology to provide auditory at-home support for students such as those who speak a language other than the language being used for teaching. Text-to-speech technology was used to read the content on a screen, allowing students to obtain both visual and auditory information simultaneously.

Analysis of learning behavior. In this dimension, we reviewed recognition technologies that were used for the analysis of learning behavior. That is, recognition technologies detected learning behavior, analyzed it, and then provided corrective feedback and / or guided learners (Barmaki et al., 2018). Gesture recognition technology received a significant amount of attention in the reviewed studies. For example, this technology was employed to assist with learning and writing Chinese characters (Hao et al., 2010; Hong et al., 2017), to improve fine motor skills and recognition in children with autism spectrum disorder (Cai et al., 2018), pottery wheel-throwing (Glushkova & Manitsaris, 2018), and English as a foreign language (Yang & Liao, 2014). Students in these studies used gestures to interact with digital objects, e.g., students used a potter's wheel, and the system provided feedback on how well the throwing moves were. The reviewed studies demonstrated the positive effects of gesture-based interactions on learning.

Body position and movement recognition was the second most popular technology for analysis of learning behavior. In the reviewed studies, students used various postures or movements to interact with the system, and it could detect these postures or movements and give feedback to the learners. For example, students in Wang et al. (2019) learning English as a foreign language practiced newly learned vocabulary using specific postures or movements to demonstrate the meaning of specific words.

Physical activity recognition technology was used in Lindberg et al. (2016). They developed an exercise game in which wearable devices could recognize students' physical movements and improve their physical education environment. The sensors in smart phones and smart wrist bands could detect students' physical movements such as shaking, jumping, and spinning, as well as collect their heart rate data.

Eye-tracking technology is able to reveal which information students are paying attention to and how much attention they are giving that information. Eye-tracking technology was employed in Huang et al. (2015) because students were exposed to multimedia learning content including three different types of visual sensory information (i.e., a video of the instructor, slides of the lecture, and lecture captions).

Identifying students' learning status. In this dimension, recognition technologies were reviewed in terms of their ability to identify the learning status of students. Emotion recognition was the most frequently used technology in this dimension. This technology can be used to recognize the learning status of students based on facial expressions or vocal intonations and provides appropriate help and guidance based on a learner's emotional state (Bahreini et al., 2016; Chickerur et al., 2015; Lin et al., 2016). For example, scholars used emotion recognition technology to improve communication skills (Bahreini et al., 2016), to predict students' emotional states (Daouas et al., 2018), to reduce second-language speaking anxiety (Chen & Lee, 2011), to assess the effects of different types of multimedia materials on emotions associated with learning as well as learning performance (Chen & Wang, 2011), and to improve learning effects in students (Lin et al., 2016). Only one study employed attention recognition technology. Chen et al. (2017) developed an attention-aware system to assess students' attention levels. The system could identify high and low attention levels in students in an autonomous e-learning environment.

3.3. Scheme

This section is divided into participants, data collection methods, research design, and course/discipline sub-sections. In the following paragraphs, each of them is discussed individually.

Participants. Among the 72 studies found, 40.3% (n = 29) of the studies involved students from universities, e.g., undergraduate students, graduate students, and doctoral students (see Appendix III). 29.2% (n = 21) of the studies were conducted among primary school students, where participants in primary school, elementary school, and children below sixth grade were classified as primary school students. 13.9% (n = 10) of these studies involved high school and middle school students, mainly including high school students, vocational school students, and middle (or secondary) school students. 9.7% (n = 7) of the studies involved other participants, e.g., employees or experts. In 16.6% (n = 12) of the studies, the participant information was not specified. That is, the participants did not wish to reveal their demographic information. 1.4% (n = 1) included pre-school students.

Data collection methods. The researchers used questionnaires/scales (69.4%, n = 50), tests (47.2%, n = 34), interviews (34.7%, n = 25), observation (16.7%, n = 12), and focus groups (5.6%, n = 4) to collect research data (see Appendix IV). In addition, some researchers used identification technology systems to collect data. For example, Chen and Wang (2011) used the emWave system to measure the emotional state of learners. Shadiev, Huang and Hwang (2017) used NeuroSky MindWave™ to measure student attention and degree of meditation.

Research design. Among the studies found, 48.6% (n = 35) carried out experiments and set up groups for comparison, followed by not specified (34.7%, n = 25), within-subjects designs (5.6%, n = 4) and mixed-methods designs (4.2%, n = 3) (see Appendix V). The least frequently designs were single group quasi-experiment (1.4%, n = 1), single subject research designs (1.4%, n = 1), design-based research approaches (1.4%, n = 1), and interdisciplinary design approaches (1.4%, n = 1).

Course/Discipline. We listed the courses/disciplines in which recognition technologies were used (see Appendix VI). The researchers applied recognition technologies to language learning (20.8%, n = 15), mathematics (15.3%, n = 11), science education (7%, n = 5), cultural learning (5.6%, n = 4), history (5.6%, n = 4), and pre-school education (1.4%, n = 1). In addition to the above, the researchers also applied recognition technologies to other courses/disciplines. For example, physics, chemistry, PE, art, writing, dental curricula, robot programming, etc.

3.4. Advantages

The advantages of recognition technology were reported in terms of the following categories: audio, non-text image, text image, moving image, physiological signals, enhanced human-computer interaction, helping teachers more accurately understand the learning status of students, increasing student motivation and engagement, and promoting understanding during the learning process.

Audio. Biometric information recognition was found to be beneficial to learning in the reviewed studies. For example, Kaklauskas et al. (2010) concluded that a self-assessment system was superior to traditional systems because of the use of a biometric voice stress analysis. Scholars also proved the advantages of speech recognition technology to improve language learning. For example, Ahn and Lee (2016) used automatic speech recognition to improve speaking in EFL learners, where the speech recognition technology gave feedback in response to the speech input of students, and ensuing interaction helped the students interact socially and in turn stimulated their language learning. In another study, Shadiev et al. (2017) provided texts generated by speech-to-text recognition to non-native English speakers in English lectures. The texts helped students understand the lecture content.

Non-Text image. In terms of advantages of face recognition technology, according to Yueh et al. (2014), this technology can be used “to improve the learning involvement, teaching effectiveness, and quality of interaction in the context of distance education” (p. 191). The application of face recognition technology resolved earlier difficulties that the students had experienced related to seeing and recognizing one another during online discussions.

According to scholars, image-based AR technology helps students learn in real-world contexts (Hwang et al., 2018; Liou et al., 2017; Shadiev et al., 2020b). This technology recognizes objects in the real world and then creates and displays virtual information on the screen, where, for example, image-based AR recognized learning targets, and relevant questions are presented to guide students to observe targets and locate supplementary materials (Hwang et al., 2018).

Advantages of sign recognition technology were also reported. In a study by Ditcharoen et al. (2010), a Thai sign language to Thai machine translation system helped deaf students learn Thai and reduced their dependence on teachers. The system was conducive to improving the interest and motivation of these students, as well as their engagement in learning.

Text image. Two studies related to handwriting recognition revealed that this technology helped students' learn handwriting and gave learners feedback based on the handwriting recognition results (Hsiao et al., 2015; Yanikoglu et al., 2017). For example, Yanikoglu et al. (2017) found that handwriting recognition technology helped learners with handwriting exercises by detecting areas where the students needed to improve their handwriting, and Hsiao et al. (2015) demonstrated that handwriting recognition technology improved students' ability to write Chinese characters and understand the spatial structure of the characters. In terms of math expression recognition technology, according to Pacheco-Venegas et al. (2015), this tool facilitated user-computer interaction, where users could use their handwriting to interact with computers.

Moving image. Advantages of recognition technologies with respect to moving images were also reported in the reviewed articles. For example, scholars argued that gesture recognition can be used to achieve intuitive interaction. By combining augmented reality and computer vision, Yang and Liao (2014) developed a virtual English classroom in which students interacted with virtual objects through gestures. The scholars also suggested that hand movement recognition technology can be used to recognize gestures or hand movements made by students, which makes the software more interactive. Wolski and Jagodzinsk (2019) developed a virtual chemical laboratory based on a hand motion recognition system that achieved good results in a chemistry education context. In the virtual laboratory, students could simulate chemical experiments using gestures.

Physiological signals. Attention, emotion, and physical activity recognition technology based on physiological signals also has advantages. For example, in the learning process, attention recognition technology helps instructors evaluate students' attention levels; emotion recognition technology can be used to help teachers understand changes in students' emotions, and physical activity recognition technology can be used to recognize students' level of physical activity so teaching strategies can be adjusted based on levels and changes in the attention, emotions and physical activity levels of students in order to improve learning outcomes. For example, in Chen et al. (2017), the recognition system identified high and low attention levels in students in an autonomous e-learning environment and notified the instructor so that appropriate interventions could be made. The results of the reviewed studies showed that the learning process was more efficient when recognition technologies were introduced in the learning process (Chen et al., 2017; Chen & Wang, 2011; Lindberg et al., 2016).

Enhanced human-computer interaction. In the reviewed studies, students interacted with a system through recognition technologies and received feedback from the system. The results showed that human-computer interaction was carried out through voice (Ahn & Lee, 2016) or face (Yueh et al., 2014) interfaces. In addition, students interacted with objects in the real world (Hwang et al., 2018), digital objects in virtual environments (Wolski & Jagodzinsk, 2019; Yang & Liao, 2014), or created digital content (Hsiao et al., 2015; Pacheco-Venegas et al., 2015; Yanikoglu et al., 2017) by interacting with computers. This human-computer interaction was beneficial for improving various learning outcomes, e.g., language learning skills (Ahn & Lee, 2016; Yang & Liao, 2014), socialization and online learning (Yueh et al., 2014), procedural skills in science experiments (Wolski & Jagodzinsk, 2019), and character or mathematic formula writing skills (Hsiao et al., 2015; Pacheco-Venegas et al., 2015; Yanikoglu et al., 2017). For example, Ahn and Lee (2016) claimed that recognition technology provided feedback to student speech input, fostered the students to interact more and stimulated their learning. According to Hwang et al. (2018), image-based AR technology helped students learn in real-world contexts. Image-based AR recognized learning targets, where relevant questions were presented to guide students to observe targets and locate supplementary materials. In Yanikoglu et al. (2017), handwriting recognition technology helped with handwriting exercises by detecting areas where students needed to improve their handwriting. Hsiao et al. (2015) found that their system improved learners' ability to write characters and understand the spatial structure of the characters.

Helping teachers more accurately understand the learning status of students. Recognition technologies in this category could be used to help instructors determine the learning status of students with better accuracy. For example, recognition technologies were useful for instructors to evaluate attention levels in students (Chen et al., 2017) or understand changes in their emotions (Chen & Wang, 2011) while they were learning. This made it possible for the teachers to adjust their teaching strategies to respond to changes in attention or emotion in an effort to improve student learning performance.

Increase student motivation and engagement. Scholars in the reviewed studies reported that applying recognition technologies to the learning process was useful in terms of increasing student motivation and engagement. For example, Shadiev, Hwang and Liu (2018) suggested that recognition technology provides feedback on student physical activity levels, where students can compare their levels to the goals set by the instructor. If physical activity level contributed to accomplishing a goal, then the students were motivated to continue the physical activity; otherwise, feedback warned them that more effort was needed. The results of the review showed that learning was more efficient and that the system was conducive to improving interest and motivation, as well as learning engagement (Ditcharoen et al., 2010; Lindberg et al., 2016; Shadiev et al., 2018).

Promote students' understanding during learning. Recognition technologies were also found to be useful in terms of promoting student understanding during the learning process. For example, Shadiev et al. (2017) argued that those students for whom the main communication language at an academic event was not their first language were not able to fully understand the communicated content at the event. In such cases, recognition technologies could be used to assist their understanding of communicated content. Shadiev et al. (2017) provided texts generated by speech-to-text recognition to non-native English speakers in English lectures. The texts helped the students understand the lecture content.

3.5. Disadvantages

There are some limitations to recognition technologies for instruction or learning that were reported in the reviewed articles. The main limitation is the accuracy of the recognition technology. For example, some scholars reported that the recognition process was not 100% accurate (Yanikoglu et al., 2017) and that there were miss-detections and false alarms in the recognition process (Hwang et al., 2018; Lindberg et al., 2016; Pacheco-Venegas et al., 2015; Yang & Liao, 2014). One reason for this issue was the fact that the recognition technology was not mature enough (Shadiev et al., 2017). Another reason was because the stability and correctness of recognition technology can be affected by the characteristics of the signals it receives; e.g., they are not clear enough or are mixed with other unrelated signals (Hwang et al., 2018). The problem associated with the accuracy of the recognition technology can lead to distraction (Yang & Liao, 2014) or inaccurate feedback from the system (Yanikoglu et al., 2017). Therefore, recognition accuracy needs to be improved (Hsiao et al., 2015).

To address these limitations, perhaps the use of high-quality recognition equipment can improve the accuracy of recognition technology. In addition, the development of technology is now very rapid, so some new technologies may be helpful with regard to improving the accuracy of recognition technology. Also, the correct use of recognition equipment can improve the accuracy of recognition technology. Furthermore, it is possible to use multi-recognition strategies as well as adjusting the calibration of devices by using feedback or biofeedback in order to improve accuracy rates.

Some of the scholars mentioned other limitations. For example, students need to learn a new interface for recognition technologies, e.g., a gesture interface instead of using a mouse and keyboard (Yang & Liao, 2014), or wearable recognition technologies were not easy to use; e.g., headsets that detect brain waves were difficult to keep on the heads of the participants because they could slip from their position (Shadiev et al., 2017). Thus, the ease of use of recognition equipment needs to be taken into consideration (Shadiev et al., 2017; Yang & Liao, 2014). The novelty effect, i.e., the short-term improvement in learning outcomes that comes from increased interest in a new technology, was also mentioned in the literature (Shadiev et al., 2018). One may argue that other disadvantages of recognition technologies such as cost, privacy, security, and ethical issues need to be reported. For example, some recognition technologies may be too expensive and therefore inaccessible to many educators, researchers, and students. However, according to our review, no results related to any of these issues were reported in the reviewed articles.

4. Discussion and conclusion

Recognition technologies have great potential in education (Shadiev et al., 2014). However, to the best of our knowledge, the reviewed studies related to one specific recognition technology only. Therefore, there was a need for a review study that summarized all existing recognition technologies and their applications to the educational process. This study was conducted to address this gap in the literature. We reviewed seventy-two articles, and eighteen recognition technologies were identified. The following main aspects of such technologies were discussed with respect to identified technologies: (1) the recognition process, (2) applications of recognition technologies, (3) the scheme applied in the reviewed studies, such as the participants, data collection, and research design, and (4) the advantages and disadvantages of recognition technologies. Because applying recognition technologies in education is a very promising research direction, our review study may be useful to educators and researchers, especially for those with non-technical backgrounds. For example, our results can help educators and researchers understand what recognition technologies currently exist, the nature of the recognition process, the schemes and applications of recognition technologies to the field of education, and their advantages and disadvantages. Therefore, our results may be useful for educators and researchers to learn more about recognition technologies and their applications to education and may provide ideas for future research on and development of the use of recognition technologies in the field of education.

4.1. The recognition process in recognition technologies

Educators and researchers can use various recognition technologies for learning and instruction (Shadiev et al., 2014). We identified eighteen recognition technologies and grouped them into audio, non-text image, text image, moving image, physiological signals, and other categories based on the input. Audio recognition technologies receive input from the voice of the student or instructor (e.g., a student pronounces a word). Therefore, this technology can be used by educators to design learning activities intended to facilitate student pronunciation in language classrooms (Ahn et al., 2016; van Doremalen et al., 2016). Non-text image recognition technologies work based on various non-text images (e.g., objects in a classroom or student faces). For example, facial recognition technology can be used to support affective learning by identifying student emotions and providing relevant feedback (Bahreini et al., 2016), e.g., cheering up a student if the emotion is not positive. Image recognition technology can be used during the learning process to capture and recognize target objects around students and to augment with additional learning content (Hwang et al., 2018; Liou et al., 2017). Text image recognition technologies capture text features in images (e.g., mathematical formulas, letters or characters). This technology can be useful in assisting students with improving their handwriting or drawing skills (Hsiao et al., 2015; Pacheco-Venegas et al., 2015). Moving image technologies constitute various input data such as gestures, e.g., a student draws a circle to activate a context menu (Holmes et al., 2018). Educators may consider this type to recognition technology to facilitate embodied learning (Barmaki et al., 2018; Yang & Liao, 2014). Physiological signals such as the electrical activity in the brain or changes in heart rate can be also measured by recognition technologies. Educators may consider them to measure learning attention, emotions, or physical activity, for example, when student learning attention decreases (Chen et al., 2017), or students become less physically active (Lindberg et al., 2016), educators may want to intervene in order to increase attention or physical activity. Other input data such as haptic (e.g., signals generated through bodily movement by means of physical contact with devices such as joysticks and other haptics) and touch (e.g., a student touches a screen on a device) can be received and processed using recognition technologies as well. Haptic recognition technologies can be applied to science education, where students have tactile experience and feedback that creates a sense of touch through vibrations, motion, or other forces, e.g., learning about mechanics (Neri et al., 2018; Wiebe et al., 2009). Educators may consider incorporating touch recognition technology for students to learn how to interact and collaborate during problem solving tasks (Higgins et al., 2012; Mercier et al., 2017).

In addition, the review results showed that the recognition methods used for the same recognition technologies can be different. For example, for gesture recognition, either Kinect (Lai et al., 2018) or the Leap Motion Controller (Cai et al., 2018) can be used. For emotion recognition, heart rate data (Chen & Wang, 2011), facial expressions, vocal intonations (Bahreini et al., 2016), or galvanic skin response (Lin et al., 2016) can be used. Therefore, educators and researchers may select appropriate recognition technologies to use for their teaching and learning processes based on their input. Furthermore, they may consider measuring the same research variable using more than one recognition technology. In this case, the reliability and validity of measurements will be stronger.

Another issue to consider is the applicability of the research results to other educational settings. That is, scholars have reported results that are quite encouraging on most occasions, but how their approach to using specific recognition technologies will fit different educational settings in terms of age, gender, culture, location, etc. requires further exploration in future studies. Educators and researchers may also want to think about the datasets used in their research. In fact, they are all different in most studies because of the lack of a standard database. For this reason, it is difficult to make any comparison of the results with other similar methodologies under the same conditions. Therefore, a standard dataset that reflects global probability is a sorely needed resource for research in this field.

4.2. The application of recognition technologies

Our results suggest that educators and researchers may want to consider applying recognition technologies to educational contexts with respect to three categories: as a tool or method to assist learning, to analyze learning behavior, or to identify the learning status of students. In terms of the first category, recognition technologies can be applied to support learning and instruction. The results of the majority of the studies in this review showed numerous educational benefits related to the use of recognition technologies. However, it should be noted that there are obvious tendencies in the use of recognition technologies in terms of the features used to assist specific learning activities and subjects. For example, speech-based recognition technologies were applied mostly to language learning contexts, e.g., to practice speaking skills in foreign or second languages (Ahn & Lee, 2016; Mroz, 2018; Shadiev et al., 2014; van Doremalen et al., 2016), and technologies based on haptic input were used mostly in science education, e.g., students learned concepts related to the field of mechanics by manipulating virtual objects with tactile-sensory tools (Neri et al., 2018; Hamza-Lup et al., 2010; Schönborn et al., 2011). Touch recognition technology was applied to support collaborative learning activities among students in most studies because it features large multi-touch operated desktops (Higgins et al., 2012; Hwang et al., 2015; Mercier et al., 2017). It was also noted that students used touch recognition technology to learn mathematics or history in most of the reviewed studies involving this type of technology.

In terms of the second category, recognition technologies can be used to analyze the learning behavior of students. This will help instructors determine the learning behavior of students and apply appropriate teaching strategies in order to improve learning outcomes. Scholars in most of the studies agreed that recognition technologies are valid, reliable tools for analyzing learning behavior. Educators and researchers should however be aware that a tendency was observed indicating that recognition technologies in this category were used for analyzing student body movements such as handwriting, gestures, or postures. Physical education and language learning were the two subjects in which recognition technologies for learning behavior were employed the most (Hao et al., 2010; Hong et al., 2017; Lindberg et al., 2016).

In terms of the third category, scholars in the reviewed studies used recognition technologies to identify the learning status of students, which is important information. If the learning status is not appropriate, i.e., students are inattentive, anxious, or bored, then the instructors need to intervene. A tendency was observed toward an increase in the number of studies focusing on detecting student emotions and attention levels. The data for emotion recognition generally comprised variations in student facial expressions, voice, and heart rate (Bahreini et al., 2016; Chen et al., 2017; Chickerur et al., 2015).

Our results showed several tendencies in the usage of recognition technologies with respect to the three categories discussed above. This can be explained by the capacity and features of recognition technologies to detect and process inputted information. For example, speech recognition technology receives input from the voice, and therefore, it was employed in a context language acquisition context, specifically to improve speaking skills. Nevertheless, there are other educational contexts (i.e., not following such tendencies) to which recognition technologies can be applied, and educators and researchers may want to think about innovative ways to do so. For example, speech recognition technology can be used in different areas other than the language learning domain. One possible application is using speech recognition technology for transcription of a lecturer's spoken content during lectures. This approach could be applied to any domain knowledge rather than only to language learning. Transcripts can be used by students during lectures to take notes, to enhance their understanding of lecture content when some important information is missed or misheard, and after lectures to complete homework assignments.

Another issue is that the research in this field is relatively thin. That is, in one decade, there were only seventy-two studies published. Therefore, more research should be done, especially with respect to those under-used recognition technologies in educational contexts such as attention recognition, biometric information recognition, and math expression recognition.

4.3. Schemes of the reviewed studies, such as the participants, data collection, and research design

In the reviewed studies, we found that the largest proportion of participants comprised university and primary school students. This finding suggests that educators and scholars in most of the reviewed studies used recognition technologies to support either higher or primary education. This finding demonstrates that scholars did not involve other subjects (e.g., middle/secondary school students, in-service teachers, or participants involved in on-the-job training). We also found that, in almost one fifth of the studies, the participants' demographic information was not shown. Therefore, it is suggested that, in the future, studies can be carried out in educational settings involving these under-researched groups of learners. This would help enrich the knowledge base related to such groups. We also suggest that future studies provide the participants' demographic information because it is important for the generalizability of the results, as well as for the internal and external validity of the cohort design.

Many of the researchers used questionnaires to collect the participants' views and perceptions, such as their degree of satisfaction with using the technology. Researchers also often used tests to measure learning outcomes. Questionnaires are very popular instruments to collect subjective data since the data can be collected easily from a large population. On the other hand, tests are very reliable instruments and are used for experimental studies to show the effectiveness of interventions. This finding suggests that other instruments were under-used by the researchers. Therefore, it is suggested that educators and researchers consider using other instruments as well.

With respect to research design, nearly half of the studies set up group comparisons. Some studies compared the learning outcomes of the experimental group with those of a control group. This indicates that the most common experimental method in the reviewed studies was a group comparison. One major concern with this finding is the lack of any comparative analysis with other benchmark works in the field (i.e., other similar recognition technologies). In future studies, researchers may want to consider carrying out such comparative analyses because they may show how good/bad the results are compared to those when other similar recognition technologies are used. Our findings also show that, in previous studies, other research designs were considered less. For example, a design-based research approach was employed in only one study (Yanikoglu et al., 2017). In addition, descriptive, explanatory, or correlational research designs were not used at all. Therefore, we suggest that, in the future, scholars may want to use other designs in their research as well. According to our results, the research design was not specified in twenty-three studies, so scholars need to explicitly report such important information in the future.

We found that research on recognition technologies can be divided into two main streams: development (i.e., some scholars developed and evaluated recognition technologies in their research) and application (i.e., some scholars applied existing recognition technologies to support learning and instruction). In the former, the researchers introduced the recognition technology system they developed, the framework and model used, and then evaluated the technology. For example, Bahreini et al. (2016) developed emotional recognition software and validated the recognition results. Daouas et al. (2018) proposed an emotionally intelligent e-learning system, introduced its architecture and the model used, and conducted a validation test. In the latter, the researchers used existing recognition technology tools and applied them to support learning and instruction. For example, Hao et al. (2010) and Hong et al. (2017) designed games based on gesture recognition to help learners learn Chinese. Mercier et al. (2017) studied cooperative learning with a multi-touch technology. Based on this finding, we suggest that various recognition technologies can be developed and tested based on different models and frameworks, or scholars may want to use existing recognition tools to support teaching and learning processes.

4.4. Advantages and disadvantages of recognition technologies

It is important for educators and researchers to know the advantages and disadvantages of recognition technologies and their applications to education. Such knowledge will enable stakeholders to use known advantages and to minimize known disadvantages in order for learning and instruction to be successful.

Researchers have demonstrated that most recognition tools have the potential to support the teaching and learning process. Nevertheless, the results of this review study demonstrated that recognition technology accuracy is a common problem in studies on recognition technologies (Hsiao et al., 2015; Lindberg et al., 2016; Pacheco-Venegas et al., 2015; Yang & Liao, 2014). Therefore, it is suggested that, in the future, improving the accuracy of recognition technology remains an important issue.

Another issue in recognition technology-related research arises from the fact that some technologies are intrusive or uncomfortable to use. For example, some students had a hard time keeping headsets used in recognition technology on their heads (Shadiev et al., 2017). This issue implies that it will be impossible to apply some of recognition technologies in real-life situations when the input data gas to be received by technologies during typical learning or working processes. Therefore, it should be noted that most of reviewed studies tested the effectiveness of recognition technology applications in laboratory settings, so how student learning outcomes can be improved in the real world is still not known. In this regard, we can say that application of recognition technologies to assist learning and instruction is still in the early stages of research. Therefore, more studies are required, especially in real world settings where learning usually takes place (e.g., real classrooms, schools, communities, or homes), and how to make technologies less intrusive or make students less uncomfortable when using technologies needs to be considered in the future. Furthermore, such issues as the novelty effect, cost, privacy, security, and ethical issues while using recognition technologies need to be considered and reported in future studies.

4.5. Trends

The results demonstrated that trend toward publication is increasing. This suggests that recognition technologies are getting better in terms of their accuracy rates, ease of use, and usefulness for learning and instruction. As a result, scholars are beginning to pay more attention to such technologies and their applications to education. The results also showed that some technologies have received little attention in research compared to other technologies. For example, biometric information recognition or sign recognition were the least frequently used, and speech recognition or touch recognition were the most frequently used. This finding suggests that educators who are planning to apply recognition technologies and seek information about their desired applications to education may find very little about such recognition technologies such as biometric information recognition or sign recognition. Therefore, more research related to applications of such technologies in education is needed.

Future studies may consider the extension of recognition technologies with other research tools. For example, recognition technologies can be combined with artificial intelligence to facilitate learning. Artificial intelligence (AI) contains multiple fields and content, and recognition technologies will ultimately be designed and developed for the artificial intelligence educational field. Therefore, multi-sensory recognition schemes will be a trend to improve accuracy in the future. In addition, fueled by AI technology, recognition technology will become more accurate and smarter. In addition, the integration of virtual reality and recognition technology such as image-based AR is an important research issue worth investigating. In the future, researchers may also want to consider combining multiple recognition technologies to enhance learning. For example, it is possible to combine face and speech recognition in order to receive image and audio input from recognition systems that will improve accuracy rates and also make such systems more reliable. Finally, future recognition technologies can identify intrinsic mental and psychological activities, such as imagination, attention, and creativity, more accurately, and perhaps, future recognition machines can be trained to recognize targets more accurately.

4.6. Research limitations

Several limitations of this study need to be acknowledged and addressed in the future. We reviewed eighteen different recognition technologies in the application domains but not the technologies themselves. Therefore, scholars may wish to address this limitation in future studies.

In this review study, we only included articles that our search yielded based on specific search terms. This approach could have excluded some articles on certain recognition technologies if their author/s did not use any of these terms (e.g., recognition). For example, our search yielded, and we reviewed, only three studies related to image-based AR (see Appendix I). However, one may argue that there are potentially many more related studies. Therefore, future studies may consider including other search terms in addition to those that were used in this study, e.g., future studies may use such search terms as “augmented reality” in order to identify more studies related to image-based AR.

Another limitation relates to certain aspects of recognition technology applications intended to assist learning and instruction, such as its efficiency, accuracy, and cost. It would be very useful for educators and researchers to know how efficient, accurate, and costly certain recognition technology is and the relationship between such aspects and technology application for effective and successful learning and instruction, e.g., how accurate a recognition technology is and how its accuracy impacts learning and instruction outcomes. Exploring these aspects of recognition technologies were out of scope of the present research and should be considered by scholars in future review studies.

References

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

Ardito, C., Lanzilotti, R., Costabile, M. F., & Desolda, G. (2013). Integrating traditional learning and games on large displays: An Experimental study. *Educational Technology & Society*, *16*(1), 44–56.

Bahreini, K., Nadolski, R., & Westera, W. (2016). Towards multimodal emotion recognition in e-learning environments. *Interactive Learning Environments, 24*(3), 590-605.

Barmaki, R., & Hughes, C. E. (2018). Embodiment analytics of practicing teachers in a virtual immersive environment. *Journal of Computer Assisted Learning*, *34*(4), 387-396.

Bause, I. M., Brich, I. R., Wesslein, A. K., & Hesse, F. W. (2018). Using technological functions on a multi-touch table and their affordances to counteract biases and foster collaborative problem solving. *International journal of computer-supported collaborative learning*, *13*(1), 7-33.

Cai, S., Zhu, G., Wu, Y. T., Liu, E., & Hu, X. (2018). A Case study of gesture-based games in enhancing the fine motor skills and recognition of children with autism. *Interactive Learning Environments*, *26*(8), 1039-1052.

Calvo, M. G., & Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: An Integrative review. *Cognition and Emotion*, *30*(6), 1081-1106.

Caseiro, N., & Santos, D. (Eds.). (2018). *Smart specialization strategies and the role of entrepreneurial universities*. Hershey, PA: IGI Global.

Chen, C. M., & Lee, T. H. (2011). Emotion recognition and communication for reducing second-language speaking anxiety in a web-based one-to-one synchronous learning environment. *British Journal of Educational Technology, 42*(3), 417-440.

Chen, C. M., & Wang, H. P. (2011). Using emotion recognition technology to assess the effects of different multimedia materials on learning emotion and performance. *Library & Information Science Research, 33*(3), 244-255.

Chen, C. M., Wang, J. Y., & Yu, C. M. (2017). Assessing the attention levels of students by using a novel attention aware system based on brainwave signals. *British Journal of Educational Technology, 48*(2), 348-369.

Chickerur, S., & Joshi, K. (2015). 3D face model dataset: Automatic detection of facial expressions and emotions for educational environments. *British Journal of Educational Technology*, *46*(5), 1028-1037.

Daouas, T., & Lejmi, H. (2018). Emotions recognition in an intelligent elearning environment. *Interactive Learning Environments*, *26*(8), 991-1009.

Davis, P., Horn, M., Block, F., Phillips, B., Evans, E. M., Diamond, J., & Shen, C. (2015). "Whoa! We're going deep in the trees!": Patterns of collaboration around an interactive information visualization exhibit. *International Journal of Computer-Supported Collaborative Learning, 10*(1), 53-76.

Ditcharoen, N., Naruedomkul, K., & Cercone, N. (2010). SignMT: An Alternative language learning tool. *Computers & Education*, *55*(1), 118-130.

Dong, G., Wang, N., & Kuang, G. (2014). Sparse representation of monogenic signal: With application to target recognition in SAR images. *IEEE signal processing letters*, *21*(8), 952-956.

Duman, G., Orhon, G., & Gedik, N. (2015). Research trends in mobile assisted language learning from 2000 to 2012. *ReCALL*, *27*(2), 197-216.

Glushkova, A., & Manitsaris, S. (2018). Gesture recognition and sensorimotor learning-by-doing of motor skills in manual professions: A Case study in the wheel-throwing art of pottery. *Journal of Computer Assisted Learning*, *34*(1), 20-31.

Ghosh, D., Dube, T., & Shivaprasad, A. (2010). Script recognition—A Review. *IEEE Transactions on pattern analysis and machine intelligence*, *32*(12), 2142-2161.

Grant, M. J., & Booth, A. (2009). A Typology of reviews: An Analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal, 26*(2), 91–108.

Hamza-Lup, F. G., & Stanescu, I. A. (2010). The Haptic paradigm in education: Challenges and case studies. *The Internet and Higher Education*, *13*(1-2), 78-81.

Hao, Y., Hong, J. C., Jong, J. T., Hwang, M. Y., Su, C. Y., & Yang, J. S. (2010). Non-native Chinese language learners' attitudes towards online vision-based motion games. *British Journal of Educational Technology*, *41*(6), 1043-1053.

Higgins, S., Mercier, E., Burd, L., & Joyce-Gibbons, A. (2012). Multi-touch tables and collaborative learning. *British Journal of Educational Technology, 43*(6), 1041-1054.

Horn, M. S., Phillips, B. C., Evans, E. M., Block, F., Diamond, J., & Shen, C. (2016). Visualizing biological data in museums: Visitor learning with an interactive tree of life exhibit. *Journal of Research in Science Teaching*, *53*(6), 895-918.

Howard, A. M., Park, C. H., & Remy, S. (2012). Using haptic and auditory interaction tools to engage students with visual impairments in robot programming activities. *IEEE transactions on learning technologies*, *5*(1), 87-95.

Hsiao, H. S., Chang, C. S., Chen, C. J., Wu, C. H., & Lin, C. Y. (2015). The Influence of Chinese character handwriting diagnosis and remedial instruction system on learners of Chinese as a foreign language. *Computer Assisted Language Learning*, *28*(4), 306-324.

Hwang, G. J., Chang, S. C., Chen, P. Y., & Chen, X. Y. (2018). Effects of integrating an active learning-promoting mechanism into location-based real-world learning environments on students' learning performances and behaviors. *Educational Technology Research & Development*, *66*(2), 451-474.

Hwang, W. Y., Shadiev, R., Tseng, C. W., & Huang, Y. M. (2015). Exploring effects of multi-touch tabletop on collaborative fraction learning and the relationship of learning behavior and interaction with learning achievement. *Educational Technology & Society, 18*(4), 459-473.

Kaklauskas, A., Zavadskas, E. K., Pruskus, V., Vlasenko, A., Seniut, M., Kaklauskas, G., Matuliauskaite, A., & Gribniak, V. (2010). Biometric and intelligent self-assessment of student progress system. *Computers & Education*, *55*(2), 821-833.

Khalil, R. A., Jones, E., Babar, M. I., Jan, T., Zafar, M. H., & Alhussain, T. (2019). Speech emotion recognition using deep learning techniques: A Review. *IEEE Access*, *7*, 117327-117345.

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

Lindberg, R., Seo, J., & Laine, T. H. (2016). Enhancing physical education with exergames and wearable technology. *IEEE Transactions on Learning Technologies*, *9*(4), 328-341.

Liou, H. H., Yang, S. J., Chen, S. Y., & Tarng, W. (2017). The Influences of the 2D image-based augmented reality and virtual reality on student learning. *Journal of Educational Technology & Society*, *20*(3), 110-121.

Liu, H., & Wang, L. (2018). Gesture recognition for human-robot collaboration: A Review. *International Journal of Industrial Ergonomics*, *68*, 355-367.

Magana, A. J., & Balachandran, S. (2017). Unpacking students conceptualizations through haptic feedback. *Journal of Computer Assisted Learning*, 33(5), 513-531.

McCrocklin, S. M. (2016). Pronunciation learner autonomy: The Potential of automatic speech recognition. *System*, *57*, 25-42.

Mercier, E. M., Higgins, S. E., & Da Costa, L. (2014). Different leaders: Emergent organizational and intellectual leadership in children's collaborative learning groups. *International Journal of Computer-Supported Collaborative Learning, 9*(4), 397-432.

Mercier, E., Vourloumi, G., & Higgins, S. (2017). Student interactions and the development of ideas in multi-touch and paper-based collaborative mathematical problem solving. *British Journal of Educational Technology*, *48*(1), 162-175.

Mroz, A. (2018). Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. *Foreign Language Annals*, *51*(3), 617-637.

Neri, L., Noguez, J., Robledo-Rella, V., Escobar-Castillejos, D., & Gonzalez-Nucamendi, A. (2018). Teaching of classical mechanics concepts using visuo-haptic simulators. *Educational Technology & Society, 21*(2), 85–97.

Pacheco-Venegas, N. D., López, G., & Andrade-Aréchiga, M. (2015). Conceptualization, development and implementation of a web-based system for automatic evaluation of mathematical expressions. *Computers & Education*, *88*, 15-28.

San Diego, J. P., Cox, M. J., Quinn, B. F., Newton, J. T., Banerjee, A., & Woolford, M. (2012). Researching haptics in higher education: The Complexity of developing haptics virtual learning systems and evaluating its impact on students' learning. *Computers & Education*, *59*(1), 156-166.

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

Shadiev, R., Hwang, W. Y., & Liu, T. Y. (2018). A Study of the use of wearable devices for healthy and enjoyable English as a foreign language learning in authentic contexts. *Educational Technology & Society, 21*(4), 217–231.

Shadiev, R., Huang, Y. M., & Hwang, J. P. (2017). Investigating the effectiveness of speech-to-text recognition application on learning performance, attention, and meditation. *Educational Technology Research and Development, 65*(5), 1239-1261.

Shadiev, R., & Sintawati, W. (2020a). A Review of research on intercultural learning supported by technology. *Educational Research Review*, 100338. doi:10.1016/j.edurev.2020.100338

Shadiev, R., Wu, T. T., & Huang, Y. M. (2020b). Using image-to-text recognition technology to facilitate vocabulary acquisition in authentic contexts. *ReCALL, 32*(2), 195-212.

Shadiev, R., & Yang, M. (2020c). Review of studies on technology-enhanced language learning and teaching. *Sustainability*, *12*(2), 524. doi:10.3390/su12020524

Shih, F. Y. (2010). *Image processing and pattern recognition: Fundamentals and techniques*. Hoboken, NJ: John Wiley & Sons.

Schmitt, L. J., & Weinberger, A. (2019). Fourth graders' dyadic learning on multi-touch interfaces—Versatile effects of verbalization prompts. *Educational Technology Research and Development, 67*(3), 519-539.

van De Ville, D., Demesmaeker, R., & Preti, M. G. (2017). When Slepian meets Fiedler: Putting a focus on the graph spectrum. *IEEE Signal Processing Letters*, *24*(7), 1001-1004.

van Doremalen, J., Boves, L., Colpaert, J., Cucchiarini, C., & Strik, H. (2016). Evaluating automatic speech recognition-based language learning systems: A case study. *Computer Assisted Language Learning*, *29*(4), 833-851.

van Laere, E., & van Braak, J. (2017). Auditory support in linguistically diverse classrooms: Factors related to bilingual text-to-speech use. *Journal of Computer Assisted Learning*, *33*(1), 79-95.

Voogt, J., & Roblin, N. P. (2010). *21st century skills discussion paper.* Enschede, The Netherlands: University of Twente Press.

Wang, F., Hwang, W. Y., Li, Y. H., Chen, P. T., & Manabe, K. (2019). Collaborative kinesthetic EFL learning with collaborative total physical response. *Computer Assisted Language Learning*, *32*(7), 745-783.

Wiebe, E. N., Minogue, J., Jones, M. G., Cowley, J., & Krebs, D. (2009). Haptic feedback and students' learning about levers: Unraveling the effect of simulated touch. *Computers & Education*, *53*(3), 667-676.

Wolski, R., & Jagodziński, P. (2019). Virtual laboratory—Using a hand movement recognition system to improve the quality of chemical education. *British Journal of Educational Technology*, *50*(1), 218-231.

Yang, M. T., & Liao, W. C. (2014). Computer-assisted culture learning in an online augmented reality environment based on free-hand gesture interaction. *IEEE Transactions on Learning Technologies*, *7*(2), 107-117.

Yanikoglu, B., Gogus, A., & Inal, E. (2017). Use of handwriting recognition technologies in tablet-based learning modules for first grade education. *Educational Technology Research & Development*, *65*(5), 1369-1388.

Yueh, H. P., Lin, W., Liu, Y. L., Shoji, T., & Minoh, M. (2014). The Development of an interaction support system for international distance education. *IEEE Transactions on Learning Technologies*, *7*(2), 191-196.

Appendix

All appendices are available from the following URL: https://yadi.sk/i/R-eiKEIfnEYUIw

Greene, M. D., & Jones, W. M. (2020). Analyzing Contextual Levels and Applications of Technological Pedagogical Content Knowledge (TPACK) in English as a Second Language Subject Area: A Systematic Literature Review. *Educational Technology & Society*, *23* (4), 75–88.

Analyzing Contextual Levels and Applications of Technological Pedagogical Content Knowledge (TPACK) in English as a Second Language Subject Area: A Systematic Literature Review

Moe D. Greene* and William M. Jones

Virginia Commonwealth University, USA // Moe.greene@Fulbrightmail.org // joneswm2@vcu.edu

*Corresponding author

(Submitted April 21, 2020; Revised June 6, 2020; Accepted August 27, 2020)

ABSTRACT: Technological pedagogical content knowledge (TPACK) defines the knowledge domains required for successful technology integration. Context is identified as an important component of TPACK. The aim of this systematic literature review was to examine context levels and the application of TPACK in the area of English language teaching and learning. The empirical studies reviewed were published between 2009 and 2019. Initial database searches yielded 365 results from which 24 articles were included in the final content analysis. Analysis of the included studies revealed that classroom factors at the micro contextual level were addressed more frequently than those at the meso and macro contextual levels, which were frequently not taken into consideration in the definition and explanation of TPACK. The majority of studies used qualitative methods for data collection which were also commonly determined through self-reporting. When self-reporting is used, TPACK is exclusively viewed as knowledge that teachers possess regardless of their context. The data indicate that teacher's contextual factors such as dispositions are not always included in the operationalization of TPACK. Teachers' contextual factors highlight their perspectives and belief systems. A critical perspective of teacher's TPACK knowledge development across contexts and the roles teachers are assigned in the classroom are vital to understanding the paradigm shifts that inform teachers' practices and training.

Keywords: Technological Pedagogical Content Knowledge, Context, Contextual levels, English language, TPACK application

1. Introduction

Technological pedagogical content knowledge (TPACK) was developed in response to the absence of a theory guiding the integration of technology in education (Koehler & Mishra, 2005; Rosenberg & Koehler, 2015). Rooted in Shulman's (1986) pedagogical content knowledge (PCK) model, TPACK extends to additionally examine the knowledge of how to apply technological resources. Holistically, it considers technological knowledge (TK), pedagogical knowledge (PK) and content knowledge (CK). Rosenberg and Koehler (2015) explain that these three bodies of core knowledge coalesce to comprise technological pedagogical knowledge (TPK), pedagogical content knowledge (PCK), and technological content knowledge (TCK).

Context is identified as an important component of TPACK (Koehler & Mishra, 2008; Koh et al., 2014; Mishra & Koehler, 2006; Phillips et al., 2017; Porras-Hernandez & Salinas-Amescua, 2013; Rosenberg & Koehler, 2015; Swallow & Olofson, 2017). TPACK has been referred to as "context bound" because of how influential contextual factors (e.g., classroom design and layout, school policies, state and national technology initiatives, and teachers' technology experience) are on teachers' development of TPACK (Mishra & Koehler, 2006). The categorization and analysis of the studies in this systematic literature review was guided by a theoretical lens that views TPACK as context bound.

The conceptualizations of TPACK context vary from teachers' epistemological beliefs to classroom and institutional resources (Rosenberg & Koehler, 2015). In their analysis of 170 TPACK focused publications, Rosenberg and Koehler (2015) reported a wide variation in how context is explained and interpreted. Moreover, prior research has found context is frequently missing when TPACK is described (Porras-Hernandez & Salinas-Amescua, 2013; Rosenberg & Koehler, 2015). There is a need to investigate the extent to which context is included in studies examining TPACK, the meaning assigned to TPACK context, and contextual characteristics that enable teachers to leverage technology resources. The current literature review advances the TPACK framework by examining a more complex conception of context that includes teacher's subjective contextual factors such as dispositions to understand technology integration and TPACK knowledge construction.

2. Purpose

The current systematic literature review was conducted in consideration of previous reviews on TPACK, mainly Kelly (2010), Porras-Hernandez and Salinas-Amescua (2013), Chai, Koh and Tsai (2013b), Rosenberg and Koehler (2015), and Willermark (2018). While these reviews were considered throughout this review process, two reviews provided a framework for the current review: (a) Rosenberg and Koehler's (2015) review that examined the extent to which context is included in publications on TPACK in all subject areas, and the meaning of context when it is included, and (b) Willermark's (2018) review that examined the general characteristics of recent TPACK articles as well as the approaches used to identify teacher's application of TPACK. While these reviews did not specifically examine TPACK in English language teaching, the current review was influenced by their search strategies, search terms, limits and coding categories.

This review sharpens the focus by examining recent studies that utilized the TPACK framework in the context of English as a second language subject area. As students' demographics shift due to the increased number of English as a second language students in the United States and abroad (Clair, 1995; Dunn, 2019; Hartshorn et al., 2017; Razfar & Simon, 2011), questions of English language teachers' ability to effectively instruct these students remain. Technology integration is an important component in English language teaching. Examining the characteristics of English language teachers' TPACK, levels of context and teachers' contextual factors including dispositions could further our understanding of TPACK and its enactment in English language teaching and learning.

The review seeks to answer the following questions:

- What are the characteristics of English language teachers' TPACK in the literature on English language teaching and learning?
- What levels of context are included in the operationalization of TPACK in English language teaching and learning?
- What teachers' contextual factors, if any, does the operationalization of TPACK include?

3. Significance and definitions

Context in TPACK is limited on several levels. Porras-Hernandez and Salinas-Amescua (2013) argue that context in TPACK is referred to in a rather ambiguous manner. Contextual references in TPACK include student characteristics; classroom and institutional conditions for learning; situated teaching activities; and teacher's epistemological beliefs. Previous reviews made substantial contributions to the understanding of how context has been included in recent TPACK research, what it means when it is included and the nature of the TPACK framework (Rosenberg & Koehler, 2015; Willermark, 2018). This current review advances understanding of TPACK by including teacher's subjective contextual variables in the form of dispositions to understand characteristics of TPACK and technology integration in the context of English language teaching and learning.

Dispositions in this review are defined from a Bourdieusian perspective and are presented as a missing link between TPACK development and enactment. Bourdieu's (1977) definition of dispositions includes individual attributes, tendencies, practices and sense of the game. They are seen as the result of an individual's past experiences and experiential influences. Conceptually, dispositions are addressed in this review through two main categories: (a) teacher tendencies and beliefs (e.g., willingness to experiment with new technology, belief in the importance of technology in teaching and learning, level of comfort with technology) and (b) teacher practices (e.g., using technology to facilitate learning in the classroom, incorporating technology in planning, designing and executing lesson plans). The interaction between these categories results in a particular outcome which can explain how dispositions may shape a teacher's technological knowledge. Identifying teachers' dispositions in the context of teaching with technology provides a lens to describe and analyze the contextual factors that reciprocally affect teachers' TPACK development (Rosenberg & Koehler, 2015). In other words, the conceptualization of teachers' TPACK contextual factors is conducted in an organized and systematic way within the framework defined by dispositions.

Drawing on the framework advanced by Porras-Hernandez and Salinas-Amescua (2013), the definition of context in this review includes three levels: micro, meso, and macro. Porras-Hernandez and Salinas-Amescua (2013) describe micro level factors as those at the classroom or learning environment level, which involve in-class conditions for learning, available technologies and class norms. The meso level represents factors at the school and local community level, and is defined through the social, cultural, political, organizational, and economic conditions established there. Factors such as the availability of technology at the school level, support staff and school leadership expectations define this level. The macro level is defined through the social, political, technological, and economic conditions at the state or national level. Factors such as mandated curricular standards, initiatives related to technology development as well as national and global policies define this level. Informed by the conceptualization of context introduced by Porras-Hernandez and Salinas-Amescua (2013) and its operationalization developed by Rosenberg and Koehler (2015), levels of context in the current review are defined as follows:

- Micro: factors in the classroom affecting the development, enactment, or assessment of TPACK. This level of context includes actions and practices of teachers, classroom norms, and technology in the classroom.
- Meso: factors in the school and community affecting the development, enactment, or assessment of TPACK. This level of context includes the school system or individual schools, school culture, infrastructure related to technology, and leadership expectations.
- Macro: factors at the state, national, and global level affecting the development, enactment, or assessment of TPACK. This level of context includes larger social, political, or economic conditions of the state (or country) that shape norms as well as policies such as national curriculum standards and technology initiatives at the state or national level.

Context is important in examining and understanding how different contextual factors may impact and shape teaching practices. Trends in the literature on context levels, dispositions, and TPACK in English language teaching and learning could inform how context is addressed in programs providing professional development and training on technology integration to English language teachers. An in-depth understanding of contextualized dispositions and the characteristics of TPACK are important in facilitating the development of teachers' technological knowledge.

4. Protocol development

A review protocol was developed according to Booth's (2006) criteria to explain all aspects of the review including method, literature search strategy, sample, coding and data analysis. Booth's (2006) criteria, which are referred to as STARLITE guidelines, were followed to systematically categorize and analyze recent publications on TPACK as it relates to context and to minimize the effect of possible bias of the review process. The review protocol is described in the following sections.

4.1. Method

The method for the current review was designed according to the structure and recommendations of other systematic reviews including Chai et al. (2013b), Rosenberg and Koehler (2015), and Willermark (2018) in regard to limits, search, and coding categories. It was also informed by Booth's (2006) criteria, commonly represented with the mnemonic STARLITE (sampling strategy, type of study, approach, range of years, limits, inclusion and exclusion criteria, terms, and electronic sources).

4.2. Literature search strategy

The databases that were selected due to their coverage of TPACK in English language teaching were EBSCOhost covering Academic Search Complete, Applied Science & Technology Index and education; ERIC (Education Resources Information Center); JSTOR; and Web of Science. They include educational technology journals and teacher education journals. Electronic sources were searched using the following descriptors: Technological pedagogical content knowledge OR "TPACK" OR "TPCK" AND ESL OR EFL OR TESOL OR English as a Second Language OR English Language AND Context OR Dispositions. The search terms used were related to TPACK, English language and context. The search covered similar terms used by Chai et al. (2013b), Rosenberg and Koehler (2015) and Willermark (2018), and additional terms to reflect this review's focus on English language and context.

Broad search terms were used to get a comprehensive search result and Boolean search terms AND and OR were included to allow for as many results as possible due to variations in descriptions applied to English as a second language teaching and learning. The use of parentheses ("TPACK," "TPCK") narrowed the search to studies that included either one. The second set of terms (context, dispositions, English as a second language, TESOL, EFL, ESL, English language teaching and learning) limited the results to studies examining context and English as second language. The search was limited to articles published between 2009-2019, which overlaps with previous systematic reviews, mainly Chai et al. (2013b): 2003-2011; Rosenberg and Koehler (2015): 2005–2013; and Willermark (2018): 2011 to 2016.

4.3. Inclusion and exclusion criteria

Once all publications were collected using search strategy procedures described above, the following inclusion criteria were utilized to evaluate each research study:

- The study is written in English
- The study is peer-reviewed
- The study examines English language education (TESOL, ESL or EFL focused)
- The study is empirically based (quantitative, qualitative or mixed methods)
- The study is published between 2009–2019
- "TPCK," "TPACK," or "technological pedagogical content knowledge" included in the title, keywords, or abstract (or introduction if an abstract is not included)
- The study explicitly addresses context in the description, explanation or operationalization of TPACK
- The study explicitly states intention to explore TPACK
- The study involved pre-service or in-service teachers

The studies that are purely theoretical including systematic reviews, meta-analyses, position papers, conceptual and conference papers were excluded. To be included in the review, the article had to contain empirically based research. Title, keywords and abstract (or introduction if an abstract is not included) were manually and systematically reviewed to decide whether a study met the inclusion criteria. In case of doubt, the publication was kept for full-text reading. Primary sources deemed relevant were listed on a master reference list and a copy of each publication was obtained.

4.4. Coding procedures

The TPACK studies coded in this systematic review include references to context in the description of TPACK. Using Willermark's (2018) coding scheme, self-reporting of TPACK was divided into three subcategories: general TPACK which refers to situations where TPACK is estimated using ranking statements of TPACK constructs regardless of situation or context; specific TPACK which refers to self-reporting of teachers' actions in authentic or fictitious TPACK situations with specific scenarios; and experienced TPACK where actual experiences of conducted teaching activities involving TPACK are self-reported. Such categorization of TPACK is based on the definition of TPACK as knowledge, something teachers process such as rules and procedures of practice reported through self-reporting. Instances where TPACK is defined as competence were also reviewed for comparison. Competence is demonstrated in action and performance of TPACK. It refers to the process of transforming TPACK knowledge into instruction. Willermark (2018) argues that self-reporting of TPACK constitutes an approach that is isolated from teaching activities while TPACK as competence constitutes an application and implementation of teaching in authentic settings involving TPACK.

The categorization of publications in this review was also conducted according to Porras-Hernandez and Salinas-Amescua's (2013) levels of context as well as the Bourdieusian definition of dispositions outlined above. Table 1 summarizes data coding and data segmentation processes.

Table 1. Coding schemes for TPACK context

Variable	Description	Possible code
Definition of TPACK	TPACK as knowledge (general, specific or experienced) or TPACK as performance	1 (included) 0 (not included)
Micro	Factors at the classroom level	1 (included) 0 (not included)
Meso	Factors at the school level	1 (included) 0 (not included)
Macro	Factors at the societal level	1 (included) 0 (not included)
Teacher	Factors related to teacher including dispositions	1 (included) 0 (not included)

Only studies with an explicit focus on English language and which included context in the conceptualization of TPACK were coded for micro, meso, macro and teacher. The coding variables were found in the abstract, introduction, literature review, methods, and data analysis sections of the selected articles. Following Rosenberg and Koehler's (2015) coding strategy, the definition of TPACK is coded "1" only if TPACK is referred to as general, specific, experienced or performed (an indication of competence) knowledge. The inclusion of context levels and teachers' factors including dispositions were coded "1" each if they are included in the study and "0" if they are not present. An article could be coded "1" or "0" for multiple categories or variables. Coding information for each article selected for this review is presented in Appendix 2.

Coding procedures in this review are grounded in prior empirical and theoretical research. They were adapted from prior research including Porras-Hernandez and Salinas-Amescua (2013), Rosenberg and Koehler (2015) and Willermark (2018). To maintain the validity of research and eliminate bias, studies were coded for explicit contextual criteria. In borderline cases, an external researcher with insight into the field was consulted. Problems in coding were discussed throughout the process and the articles were reviewed by more than one person when necessary. Regular meetings were scheduled with a librarian with expertise in the field of education throughout the database search to ensure that the review and search criteria were applied consistently and exhaustively.

4.5. Study selection

A PRISMA flow diagram is provided in Figure 1 to present the flow of information throughout the systematic review process. After screening for duplicates, all articles were evaluated according to the inclusion and exclusion criteria. This phase of the screening process was straightforward as the focus was on whether there was a mention of TPACK in the title, keywords, or abstract (or introduction if an abstract is not included) and English language, context or dispositions in the rest of the article. If one or more criteria were absent or unclear, the article was considered for full review during the next screening. Studies that met the inclusion and exclusion criteria were listed on a master reference list and a copy of each publication was obtained. The second screening phase included a complete review of studies.

The search yielded a total of 365 articles (65 of them were obtained from ERIC, 46 from EBSCOhost, 55 from JSTOR and 197 From Web of Science) following a filtering procedure regarding English language and context. The number of studies that met the inclusion criteria after the final review was 24. Considering that several recent studies (Redmond & Peled, 2019; Tseng, 2018) reported that research on TPACK is limited, the number of studies that met the inclusion criteria is relatively higher than expected. TPACK has been mainly applied in discipline areas such as math and science.

PRISMA Information Flow Diagram

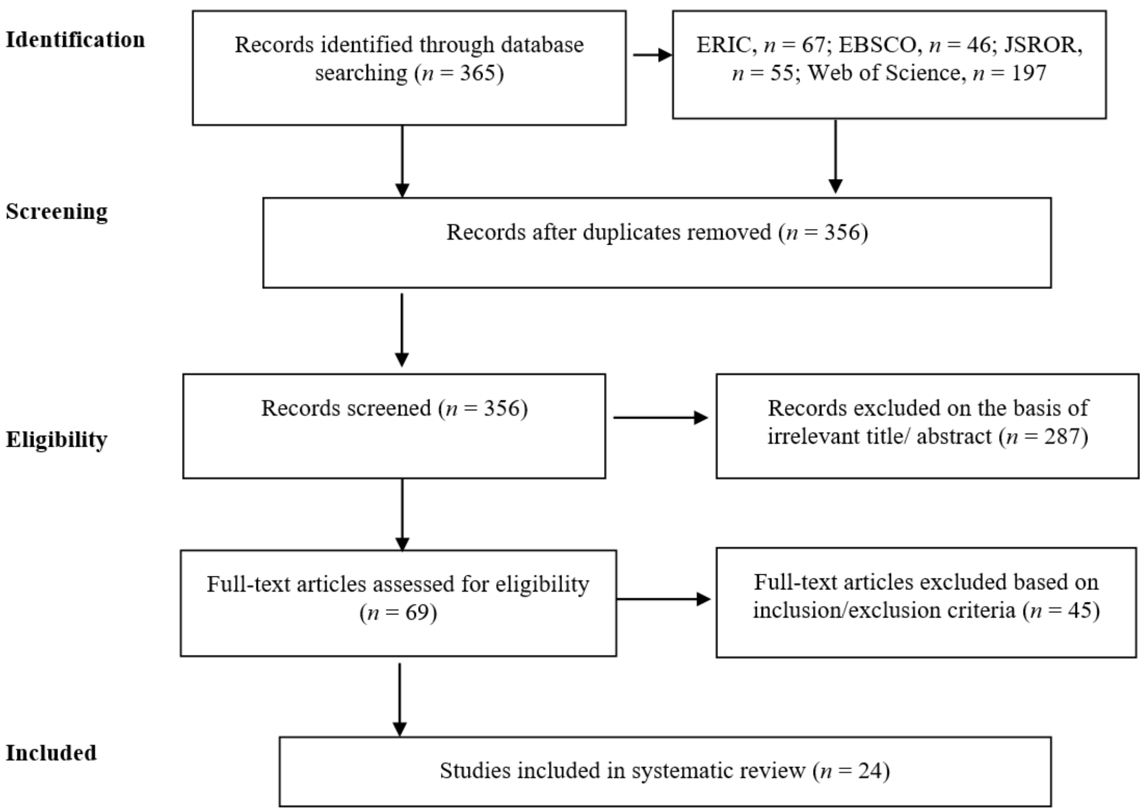


Figure 1. Flow PRISMA diagram of the screening and selection procedure

4.6. Data analysis

The "1" (included) and "0" (not included) codes were computed in the analysis of the data to determine the characteristics of TPACK in the English language teaching context. Frequencies and percentages of context levels (micro, meso, macro) and teacher factors which include dispositions were computed using the "1" (included) and "0" (not included) coding measures.

5. Results

Findings of this review are presented in four sections according to the research questions and data analysis: overview of TPACK studies, characteristics of English language teachers' TPACK, levels of context in TPACK, and English language teacher's contextual factors in TPACK

5.1. Overview TPACK studies

A total of 24 out of the initial 365 identified articles were included in this review. As shown in Figure 2, relatively few articles were included from the years of 2009 through 2015, four total. There was a clear increase of the articles that met this study's inclusion and exclusion criteria in 2016 through 2019. Most studies were published in 2018 followed by 2019.

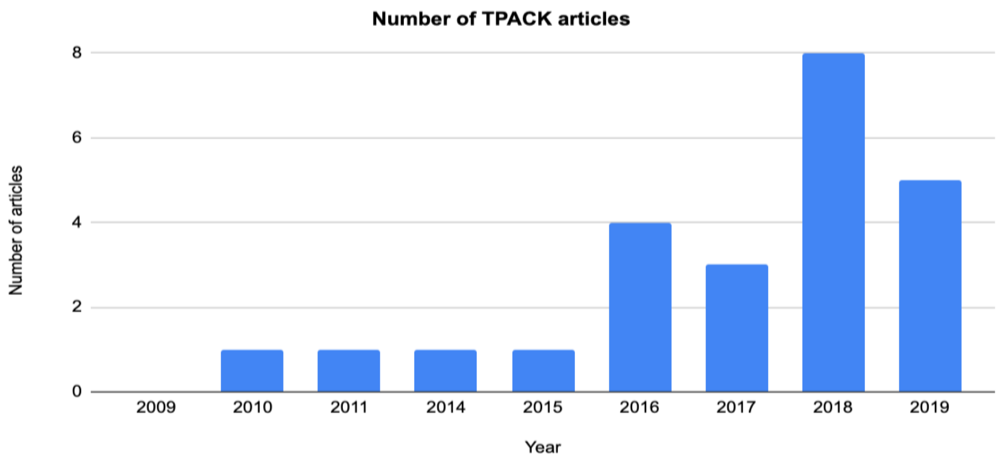


Figure 2. Distribution of the selected number of TPACK articles published between 2009 and 2019 (n = 24)

The highest number of articles, (45.83%), was found in educational technology journals (e.g., Education and Information Technologies, Australian Journal of Educational Technology, and Educational Technology & Society). 33.33% were published in language subject based journals (e.g., Journal of Asia TOEFL, Computer Assisted Language Learning and Journal of Basic Writing), 12.5% in cross discipline journals (e.g., Journal of Qualitative Research in Education and Educational Research and Reviews), and 8.33% in teacher education journals (e.g., Journal of Teacher Education and Educators, and Asia-Pacific Journal of Teacher Education). The data on the number of publications found in each journal are reported in the Figure 3 below.

As indicated in Appendix 1, eight studies—Baran et al. (2019); Baran and Uygun (2016); Chai et al. (2010); Dong et al. (2015); Joo et al. (2018); Khine et al. (2017); Redmond and Peled (2019); and Sang et al. (2016)--included in the review extended the study of TPACK to include other subject discipline areas such as social sciences education, science education, elementary mathematics education, secondary biology education, and languages. Studies with an exclusive focus on English language teaching include: Bandi-Rao and Sepp (2014); Baser et al. (2016), Bostancıoğlu and Handley (2018); Debbagh and Jones (2018); Habibi et al. (2019); Holmberg et al. (2018); Liu et al. (2019); Setiawan et al. (2018); Asık et al. (2018); Getu and Teka (2018); Tseng (2016); Tseng (2018); Tseng et al. (2011); Tseng et al. (2019); Turgut (2017a); Turgut (2017b).

The majority of selected studies used qualitative methods (37.5%; n = 9) for data collection, followed by quantitative (33.33%; n = 8) and mixed methods studies (29.16%; n = 7). The data show that pre-service teachers are the most common participants studied at 54.16% (n = 15). In-service teachers constituted 29.16% (n = 7). Both pre-service and in-service teachers were included in 8.33% (n = 2) studies. EFL students constituted 8.33%

(*n* = 2). Information about each article (including authors, year of publication, publication outlet, methods and sample groups) is included in Appendix 1.

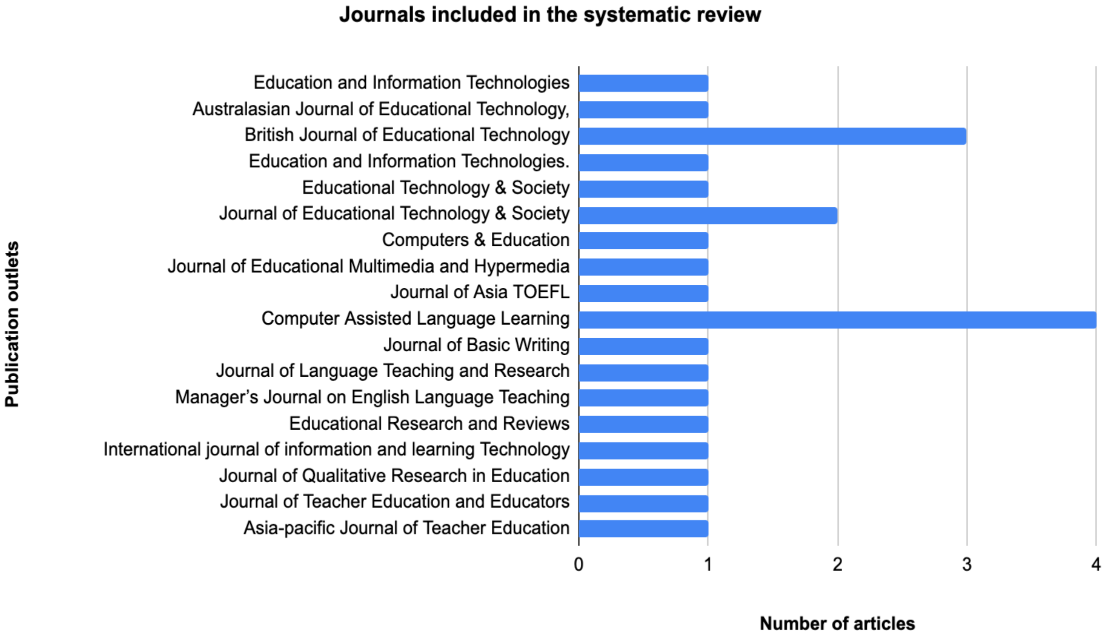


Figure 3. Distribution of the selected TPACK articles in peer reviewed journals (*n* = 18)

5.2. Characteristics of English language teachers' TPACK

Following Willermark's (2018) coding procedures, the selected articles were analyzed according to how TPACK was approached either through self-reporting or performance. The objective of such identification was to determine whether TPACK was defined as knowledge or as competence. The data show that although approaches to TPACK were often determined through a combination of instruments such as surveys and class observations to triangulate findings, TPACK was commonly determined through self-reporting at 81.82% compared to performance which constituted 18.18%.

Self-reporting was divided into three subcategories: general TPACK, specific TPACK, and experienced TPACK. The data indicate that the application of general TPACK constituted 70.73% and experienced TPACK was 29.62%. Specific TPACK was not applied in any of the articles that were reviewed. Self-reporting provided data regarding teachers' self-efficacy, beliefs and attitudes, highlighting their perspectives and belief systems about technology in teaching. Self-reporting reflects an approach where TPACK is viewed exclusively as knowledge that teachers possess regardless of their context. The majority of the studies were conducted to capture pre-service teachers' TPACK.

Several studies included in this literature review (e.g., Baser, Kopcha, & Ozden, 2016; Bostancıoğlu, & Handley, 2018; Tseng, 2016; Sang et al., 2016; Tseng, 2018) were conducted with the aim of TPACK survey validation and/or adaptation in the English language subject area. Survey completion where participants were asked to numerically rate statements on a 5-point Likert scale was the most frequently used TPACK approach. Survey questions were designed to estimate the knowledge base English language teachers are expected to possess. The most commonly used and referenced TPACK surveys were developed by Mishra and Koehler (2006); Archambault and Crippen (2009); Schmidt et al. (2009); Archambault and Barnett (2010); Chai et al. (2010); Lux et al. (2011); Sahin (2011); Chai et al. (2013a); Chai et al. (2013c); Yeh et al. (2014); and Baser et al. (2016).

Unlike general TPACK, self-reporting on experienced TPACK involved teachers' discussions or interviews on the developing and conducting lesson plans through the TPACK lens (e.g., Turgut, 2007b; Setiawan et al., 2018, and Tseng, 2018). Performance evaluation of teaching activities, where TPACK is defined as competence, usually involved class observations or pre-and post- TPACK assessments (e.g., Tseng et al., 2019; Tseng et al., 2011).

5.3. Levels of context in TPACK

The data were analyzed to determine the inclusion of context in journal articles that were selected. Studies that explicitly discussed context in reference to TPACK were coded for micro, meso and macro. The frequencies and percentages of context levels included in the definition, explanation or operationalization of TPACK are presented in Table 2.

Table 2. Context level frequencies and percentages

Context level	Frequencies	Percentages
Micro	23	46.93%
Meso	11	22.44%
Macro	15	30.61%
Total	49	100.00%

Frequencies and percentages listed in Table 2 indicate that contextual factors according to levels of context were included inconsistently among the selected journal articles. It was found that classroom factors (micro), school factors (meso) and societal factors (macro) constituted 46.93%, 22.44% and 30.16% respectively. Context level variables in the selected studies are aligned with the dimensions of the TPACK framework. Figure 4 below presents the data in regard to the inclusion of context levels.

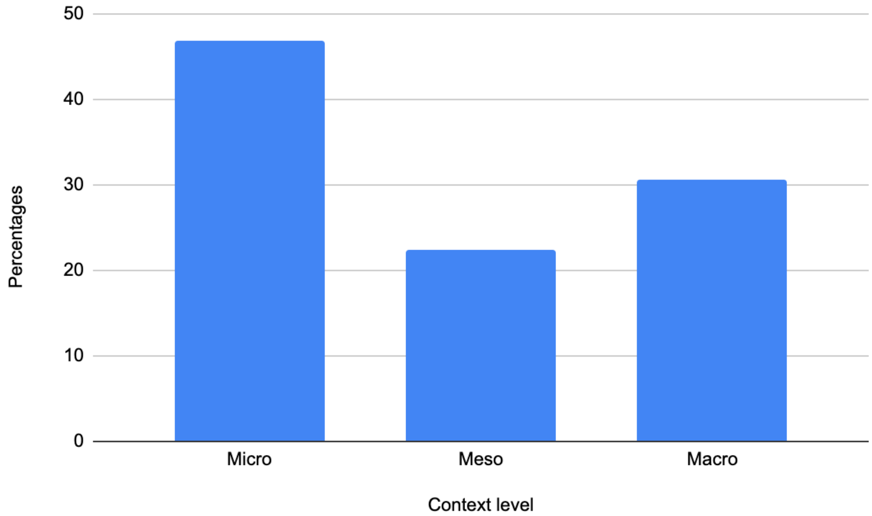


Figure 4. Context level percentages

5.4. English language teacher's contextual factors in TPACK

The operationalization of TPACK that includes dispositions as part of teachers' factors was also examined. No articles that include dispositions as defined from the Bourdieusian perspective were found. However, the data show that teachers' design dispositions in relation to TPACK were explicitly addressed in only one journal article out of 24 selected for this review. The article was published by Dong et al. (2015) who studied teachers with regards to seven factors of technological pedagogical content knowledge, their beliefs about constructivist-oriented teaching (CB) and design disposition (DD). The findings indicate that design disposition consistently predicts both pre-service and in-service teachers' TPACK. Dong et al. (2015) developed and validated a nine factor-model instrument to explore teachers' TPACK profiles and their development, highlighting the importance of design disposition for TPACK advancement.

Overall, the TPACK framework in the data gathered illustrates the interplay between levels of context (micro, meso and macro) and teachers' contextual factors. Some of the variables that were reported in the selected articles are related to teachers' background, self-efficacy, and professional roles. The data show that teachers' contextual factors are included more frequently (75%). These variables were usually framed as part of contextual factors that affect technology integration in the classroom. Moreover, a few studies emphasized student factors such as their perceptions, attitudes, and background. Student variables were presented as exerting an influence on student learning and perceptions of English language teachers' TPACK. The findings suggest that contextual factors deemed important to the conceptualization of TPACK in a study are more likely to be addressed and considered by researchers.

6. Discussion

The distribution of articles in the findings indicates that education technologists, rather than content specialists, are more interested in research on the TPACK framework in the area of English as a second language. It further suggests the need for collaboration between education technologists and content specialists. Such collaboration could shape the depth of the studies in the main English language journals such as the TESOL journal, TESOL Quarterly and the ELT journal. It can lead to a deeper consideration of how technology impacts students' learning and teacher knowledge development.

The fact that TPACK was commonly determined through self-reporting is consistent with the results from previous literature reviews including Chai et al. (2013b), Rosenberg and Koehler (2015), and Willermark (2018). Self-report measures, open-ended questionnaires, performance assessments, interviews and observation were identified to be the commonly applied measurement methods in TPACK focused studies (Koehler et al., 2011; Willermark, 2018). Willermark (2018) found that teacher self-reporting was the most frequently applied approach used to identify teacher TPACK, while performance evaluations on teaching activities were less frequent. The prevalence of self-report measures suggests a need for developing performance measures of the TPACK framework. This requires a shift in theories and approaches that guide technology integration in teaching and learning.

There is scholarly debate in the literature about the validity and applicability of the existing measurements that rely on self-reporting. Bostancıoğlu and Handley (2018) argue that several TPACK instruments (i.e., Chai et al. (2013a) and Baser et al. (2016)) appear to be influenced by dominant theories in education such as social-constructivist, communicative learning and socio-cultural theories to the exclusion of other theories of second language acquisition. They proposed an assessment of TPACK in the English language subject area which does not prescribe a particular approach or theory. There is a need for a TPACK framework that articulates constructs specific to English language teaching and learning.

The findings suggest a need to reexamine methods followed in TPACK studies. Research methods were not always mentioned in the studies selected for this review. The classification of journal articles in this review was often inferred based on other information provided. The research method classification stated in a study was used when available even if it was deemed a wrong identification for that particular study. For example, Turgut's (2017a) study was classified as mixed methods due to the fact that the data were collected through a TPACK scale (Schmidt et al., 2009) with open-ended questions and classroom observations. The author states that the qualitative data from the open-ended questions was analyzed through phenomenological data analysis. The article does not provide a detailed explanation of how the data was analyzed through a phenomenological approach. It is important to note that the results related to research methods in this review correspond with the results of Chai et al. (2013b) and differ from the results of Willermark (2018) where quantitative methods constituted the most frequently applied approach, followed by mixed methods and qualitative methods.

Consistent with the findings in previous literature reviews (e.g., Porras-Hernandez & Salinas-Amescua, 2013; Rosenberg & Koehler, 2015), the data suggest that the meaning of context has differed widely, from teachers' micro factors to institutional resources and sociocultural contextual factors. The increase in TPACK publications that take contextual factors into consideration in recent years indicate a growing interest in this area. There is a belief that teachers' knowledge and context influence how teachers incorporate technology into teaching. These findings correspond with the results of Tseng et al. (2019) that show how teachers' use of technology was moderated predominantly by teacher-centric factors at the micro level of context.

The examination of context levels and teachers' contextual factors including dispositions further our understanding of how TPACK is rooted in context. There is a need to understand both objective and subjective contextual factors that influence TPACK development and its enactment in instructional settings. Porras-Hernandez and Salinas-Amescua (2013) argue that external conditions are important elements that shape instruction at the micro level. Dispositions, on the other hand, may explain the value assigned to education, student and teacher roles, as well as access to resources and capital. Porras-Hernandez and Salinas-Amescua (2013) explained that while external contextual variables may explain the technology integration process, teacher's subjective variables can bring to light not only the teachers' technology integration process, but also the knowledge construction that takes place in a given situation.

7. Conclusion

The purpose of this review was to provide a comprehensive analysis of TPACK in English language teaching and learning along with the inclusion of contextual levels and factors in the operationalization of TPACK. The empirical studies reviewed were published between 2009 and 2019. A total of 24 out of the initial 365 identified articles were included in this review. The key findings indicate that the highest number of articles was published in 2018, followed by 2019. The majority of studies were found in educational technology journals. Qualitative methods were the most common research approach used. Context level frequencies and percentages indicate that contextual factors were included inconsistently among the selected journal articles. The data also indicate that classroom factors at the micro level were addressed more frequently. Although the data show that approaches to TPACK were often determined through a combination of instruments such as surveys and class observations, TPACK was commonly determined through self-reporting.

A wide range of TPACK assessments were identified including self-reporting and performance assessments which included class observation and interviews. The data suggest there has been a focus on creating and developing measures of TPACK that take context into consideration, a process that is still in its early stages of development. As evidenced by the increase of TPACK publications that address teachers' ability to integrate technology in instruction, there has been a shift towards the inclusion of different levels of contextual factors required for effective technology integration. The argument made was that teacher's subjective variables such as dispositions, which are absent in the studies that were reviewed, should be taken into consideration in designing and explaining TPACK applications.

This review builds on previous literature reviews (e.g., Chai et al., 2013b; Rosenberg & Koehler, 2015; Willermark, 2018) and advances the TPACK framework by including teacher's subjective contextual variables to understand technology integration and TPACK knowledge construction. Such focus has significant implications in shaping technology integration in English language teaching and in redefining TPACK to provide a multifaceted view into teachers' TPACK knowledge base and how the components of technology, pedagogy, and content are manifested in context-based practice.

While the current review provides important insight regarding the characteristics of TPACK and context in the English language subject area, there are limitations that should be noted. Only peer reviewed empirical studies were considered; there are several other publications including book chapters, conference presentations and conceptual papers that were excluded in the review. The data selection procedures produced a relatively limited number of journal articles. Following coding procedures proposed by Rosenberg and Koehler (2015), the articles were identified only if they explicitly included context. Articles that included similar, but different terms, such as situated TPACK were not included. The justification was the term context is explicitly included in the definition of TPACK by Mishra and Koehler (2006) and several other studies.

These limitations are important to recognize; however, it should be noted that this literature review is comprehensive. There are many potential areas to advance the TPACK framework, including students' learning with technology across contextual settings, addressing how the interactions of contextual factors influence and moderate TPACK knowledge development, and reviewing the current TPACK assessments in the context of English language subject area to identify issues, trends, and recommendations that could guide possible areas for future research in TPACK. Greater attention to context in research can support TPACK applications in educational technology programs that promote technology integration.

References

Archambault, L. M., & Barnett, J. H. (2010). Revisiting technological pedagogical content knowledge: Exploring the TPACK framework. *Computers & Education, 55*(4), 1656-1662.

Archambault, L. M., & Crippen, K. (2009). Examining TPACK among K-12 online distance educators in the United States. *Contemporary Issues in Technology and Teacher Education 9*(1), 71-88.

Aşık, A., İnce, B. H., & Vural, A. S. (2018). Investigating learning technology by design approach in pre-Service language teacher education: Collaborative and reflective experiences. *Journal of Qualitative Research in Education, 6*(1), 1-17.

Bandi-Rao, S., & Sepp, M. (2014). Designing a digital story assignment for basic writers using the TPCK framework. *Journal of Basic Writing, 33*(1), 103-123.

Baran, E., Canbazoglu, B. S., Albayrak, S. A., & Tondeur, J. (2019). Investigating the impact of teacher education strategies on preservice teachers' TPACK: The Impact of teacher education strategies on TPACK. *British Journal of Educational Technology, 50*(1), 357-370.

Baran, E., & Uygun, E. (2016). Putting technological, pedagogical, and content knowledge (TPACK) in action: An Integrated TPACK-design-based learning (DBL) approach. *Australasian Journal of Educational Technology, 32*(2), 47-63.

Baser, D., Kopcha, T. J., & Ozden, M. Y. (2016). Developing a technological pedagogical content knowledge (TPACK) assessment for preservice teachers learning to teach English as a foreign language. *Computer Assisted Language Learning, 29*(4), 749-764.

Booth, A. (2006). "Brimful of STARLITE": Toward standards for reporting literature searches. *Journal- Medical Library Association, 94*(4), 421-429.

Bostancıoğlu, A., & Handley, Z. (2018). Developing and validating a questionnaire for evaluating the EFL "Total PACKage": Technological pedagogical content knowledge (TPACK) for English as a Foreign Language (EFL). *Computer Assisted Language Learning, 31,* 572-598.

Bourdieu, P. (1977). *Outline of a theory of practice*. Cambridge, England: Cambridge University Press.

Chai, C. S., Chin, C. K., Koh, J. H., & Tan, C. L. (2013a). Exploring Singaporean Chinese language teachers' technological Pedagogical content knowledge and its relationship to the teachers' pedagogical beliefs. *Asia-pacific Education Researcher, 22*(4), 657-666.

Chai, C. S., Koh, J. H., & Tsai, C. C. (2010). Facilitating preservice teachers' development of Technological, Pedagogical, and Content Knowledge (TPACK). *Educational Technology & Society, 13*(4), 63-73.

Chai, C. S., Koh, J. H., & Tsai, C. C. (2013b). A Review of technological pedagogical content knowledge. *Educational Technology & Society*, *16*(2), 31–51.

Chai, C. S., Ng, E. M., Lee, W. H., Hong, H. Y., & Koh, J. H. (2013c). Validating and modeling technological pedagogical content knowledge (TPCK) framework among Asian pre- service teachers. *Australasia Journal of Educational Technology, 29*(1), 41–53.

Clair, N. (1995). Mainstream classroom teachers and ESL students. *TESOL Quarterly, 29*(1), 189-196.

Debbagh, M., & Jones, W. M. (2018). Examining English language teachers' TPACK in oral communication skills teaching. *Journal of Educational Multimedia and Hypermedia, 27*(1), 43-62.

Dong, Y., Sang, G. Y., Chai, C. S., Koh, J. H., & Tsai, C. C. (2015). Exploring the profiles and interplays of pre-service and in-service teachers' technological pedagogical content knowledge (TPACK) in China. *Educational Technology and Society, 18*(1), 158-169.

Dunn, S. C. (2019). Learning to communicate across language and culture: Demographic change, schools, and parents in adult ESL classes. *School Community Journal, 29*(1), 9-38.

Getu, T., & Teka, M. (2018). Pedagogical content knowledge (PCK) in Ethiopian secondary school teacher education practicum supervision. *Journal of Teacher Education and Educators, 7*(2), 165-181.

Habibi, A., Yusop, F. D., & Razak, R. A. (2019). The Role of TPACK in affecting pre-service language teachers' ICT integration during teaching practices: Indonesian context. *Education and Information Technologies, 25*(3), 1929-1949.

Hartshorn, K. J., Evans, N. W., Egbert, J., & Johnson, A. (2017). Discipline-specific reading expectation and challenges for ESL learners in US universities. *Reading in a Foreign Language, 29*(1), 36-60.

Holmberg, J., Fransson, G., & Fors, U. (2018). Teachers' pedagogical reasoning and reframing of practice in digital contexts. *The International Journal of Information and Learning Technology, 35*(2), 130-142.

Joo, Y. J., Park, S., & Lim, E. (2018). Factors influencing preservice teachers' intention to use technology: TPACK, teacher self-efficacy, and technology acceptance model. *Educational Technology & Society, 21*(3), 48-59.

Kelly, M. (2010). Technological pedagogical content knowledge (TPACK): A Content analysis of 2006-2009 print journal articles. In D. Gibson & B. Dodge (Eds.), *Society for Information Technology & Teacher Education International Conference* (Vol. 2010, pp. 3880–3888). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE). Retrieved from http://www.editlib.org/p/33985/

Khine, M. S., Ali, N., & Afari, E. (2017). Exploring relationships among TPACK constructs and ICT achievement among trainee teachers. *Education and Information Technologies, 22*(4),1605-1621.

Koehler, M. J., & Mishra, P. (2005). What happens when teachers design educational technology? The Development of Technological Pedagogical Content Knowledge. *Journal of Educational Computing Research, 32*(2), 131-152.

Koehler, M. J., & Mishra, P. (2008). Introducing TPCK. In AACTE Committee on Technology and Innovation (Eds.), *Handbook of Technological Pedagogical Content Knowledge (TPCK) for Educators* (pp. 3–29). New York, NY: Routledge.

Koehler, M., Shin, T., & Mishra, P. (2011). How do we measure TPACK? Let me count the ways. In R. N. Ronau, C. R. Rakes, & M. L. Niess (Eds.), *Educational technology, teacher knowledge, and classroom impact: A Research Handbook on Frameworks and Approaches* (pp. 16–31). Hershey, PA: Information Science Reference.

Koh, J. H., Chai, C. S., & Tay, L. Y. (2014). TPACK-in-Action: Unpacking the contextual influences of teachers' construction of technological pedagogical content knowledge (TPACK). *Computers & Education, 78*, 20–29.

Liu, H., Wang, L., & Koehler, M. J. (2019). Exploring the intention-behavior gap in the technology acceptance model: A Mixed-methods study in the context of foreign-language teaching in China. *British Journal of Educational Technology, 50*(5), 2536-2556.

Lux, N. J., Bangert, A. W., & Whittier, D. B. (2011). The Development of an instrument to assess preservice teacher's technological pedagogical content knowledge. *Journal of Educational Computing Research, 45*(4), 415-431.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A Framework for teacher knowledge. *Teachers College Record, 108*(6), 1017–1054.

Phillips, M., Koehler, M., Rosenberg, J., & Zunica, B. (2017). Unpacking TPACK: Reconsidering knowledge and context in teacher practice. In P. Resta & S. Smith (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference* (pp. 2422-2429). Austin, TX, United States: Association for the Advancement of Computing in Education (AACE). Retrieved from https://www.learntechlib.org/primary/p/177538/

Porras-Hernandez, L. H., & Salinas-Amescua, B. (2013). Strengthening TPACK: A Broader notion of context and the use of teacher's narratives to reveal knowledge construction. *Journal of Educational Computing Research, 48*, 223–244.

Razfar, A., & Simon, J. (2011). Course-taking patterns of Latino ESL students: Mobility and mainstreaming in urban community colleges in the United States. *TESOL Quarterly, 45*(4), 595-627.

Redmond, P., & Peled, Y. (2019). Exploring TPACK among pre-service teachers in Australia and Israel. *British Journal of Educational Technology, 50*(4), 2040-2054.

Rosenberg, J. M., & Koehler, M. J. (2015). Context and technological pedagogical content Knowledge (TPACK): A Systematic review. *Journal of Research on Technology in Education, 47*(3), 186-210.

Sahin, I. (2011). Development of survey of technological pedagogical and content knowledge (TPACK). *Turkish Online Journal of Educational Technology, 10*(1), 97-105.

Sang, G., Tondeur, J., Chai, C. S., & Dong, Y. (2016). Validation and profile of Chinese pre-service teachers' technological pedagogical content knowledge scale. *Asia-pacific Journal of Teacher Education, 44*(1), 49-65.

Schmidt, D. A., Baran, E., Thompson, A. D., Mishra, P., Koehler, M. J., & Shin, T. S. (2009). Technological pedagogical content knowledge (TPACK): The Development and validation of an assessment instrument for preservice teachers. *Journal of Research on Technology in Education, 42*(2), 123–149.

Setiawan, I., Hamra, A., Jabu, B., & Susilo, S. (2018). Exploring a teacher educator's experiences in modeling TPACK to create English language multimedia in technology courses. *Journal of Language Teaching and Research, 9*(5), 1041-1052.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.

Swallow, M. J., & Olofson, M. W. (2017). Contextual understandings in the TPACK Framework. *Journal of Research on Technology in Education, 49,* 228-244.

Tseng, J. J., Cheng, Y. S., & Lin, C. C. (2011). Unraveling in-service EFL teachers' technological pedagogical content knowledge. *Journal of Asia TEFL, 8*(2), 45-72.

Tseng, J. J. (2016). Developing an instrument for assessing technological pedagogical content knowledge as perceived by EFL students. *Computer Assisted Language Learning, 29*(2), 302-315.

Tseng, J. J. (2018). Exploring TPACK-SLA interface: Insights from the computer-enhanced classroom. *Computer Assisted Language Learning, 31*(4), 390-412.

Tseng, J. J., Cheng, Y. S., & Yeh, H. N. (2019). How pre-service English teachers enact TPACK in the context of web-conferencing teaching: A Design thinking approach. *Computers and Education, 128,* 171-182.

Turgut, Y. (2017a). An Analysis of TPACK integration into English language teacher education programs: A Comparative study of Turkey and Denmark. *Journal on English Language Teaching, 7*(4), 1-15.

Turgut, Y. (2017b). A Comparison of pre-service, in-service and formation program for teachers' perceptions of technological pedagogical content knowledge (TPACK) in English language teaching (ELT). *Educational Research and Reviews, 12*(22), 1091-1106.

Willermark, S. (2018). Technological pedagogical and content Knowledge: A Review of empirical studies published from 2011 to 2016. *Journal of Educational Computing Research, 56*(3), 315-343.

Yeh, Y., Hsu, Y., Wu, H., Hwang, F., & Lin, T. (2014). Developing and validating technological pedagogical content knowledge-practical (TPACK-practical) through the Delphi survey technique. *British Journal of Educational Technology, 45*, 707–722.

Appendix 1: Articles and codes for the publications included in the systematic review (publication outlet, research method, and population sample)

Article #	Authors & Year	Publication outlet	Research method	Population sample
TPACK studies that focused on English as a second language subject area only				
1	Bandi-Rao and Sepp (2014)	Journal of Basic Writing	Qualitative	EFL students
2	Baser, Kopcha and Ozden 2016)	Computer Assisted Language Learning	Mixed methods	Pre-service teachers
3	Bostancıoğlu and Handley (2018)	Computer Assisted Language Learning,	Mixed methods	In-service teachers
4	Debbagh and Jones (2018)	Journal of Educational Multimedia and Hypermedia	Qualitative	In-service teachers
5	Habibi, Yusop and Razak (2019)	Education and Information Technologies	Quantitative	Pre-service teachers
6	Holmberg, Fransson and Fors (2018)	International Journal of Information and Learning Technology	Qualitative	In-service teachers
7	Liu, Wang and Koehler (2019)	British Journal of Educational Technology	Mixed methods	In-service teachers
8	Setiawan, Hamra, Jabu and Susilo (2018)	Journal of Language Teaching and Research	Qualitative	In-service teachers
9	Aşık, İnce and Vural (2018)	Journal of Qualitative Research in Education	Mixed methods	Pre-service teachers
10	Teka (2018)	Journal of Teacher Education and Educators	Qualitative	Pre-service teachers
11	Tseng (2016)	Computer Assisted Language Learning	Quantitative	EFL students
12	Tseng (2018)	Computer Assisted Language Learning	Mixed methods	In-service teachers
13	Tseng, Cheng and Lin (2011)	Journal of Asia TOEFL	Qualitative	In-service teachers
14	Tseng, Cheng and Yeh (2019)	Computers & Education	Qualitative	Pre-service teachers
15	Turgut (2017 b)	Educational Research and Reviews	Mixed methods	Pre-service & In-service teachers
16	Turgut (2017 a)	Manager’s Journal on English Language Teaching	Mixed methods	Pre-service teachers
TPACK studies that included other subject discipline areas such as social sciences education, science education etc.				
17	Baran and Uygun (2016)	Australasian Journal of Educational Technology,	Qualitative	Pre-service teachers
18	Baran, Canbazoglu, Albayrak and Tondeur (2019)	British Journal of Educational Technology	Quantitative	Pre-service teachers
19	Chai, Koh and Tsai (2010).	Educational Technology & Society	Quantitative	Pre-service teachers
20	Dong, Sang, Chai, Koh and Tsai (2015)	Educational Technology & Society	Quantitative	Pre-service & In-service teachers
21	Joo, Park and Lim (2018)	Educational Technology & Society	Quantitative	Pre-service teachers
22	Khine, Ali and Afari (2017)	Education and Information Technologies	Quantitative	Pre-service teachers
23	Redmond and Peled (2019)	British Journal of Educational Technology	Qualitative	Pre-service teachers
24	Sang, Tondeur, Chai and Dong (2016)	Asia-pacific Journal of Teacher Education	Quantitative	Pre-service teachers

Appendix 2: Articles and codes for the publications included in the systematic review (TPACK variables and levels of contextual)

Article #	Authors & Year	TPACK as knowledge			TPACK as performance	TPACK context levels			Dispositions	Teacher's factors
		General	Specific	Experienced		Micro	Meso	Macro		
TACK studies that focused on English as a second language subject area only										
1	Bandi-Rao and Sepp (2014)			X		X	X	X		X
2	Baser, Kopcha and Ozden 2016)	X				X		X		X
3	Bostancıoğlu and Handley (2018)	X				X		X		X
4	Debbagh and Jones (2018)	X			X	X				X
5	Habibi, Yusop and Razak (2019)	X				X		X		X
6	Holmberg, Fransson and Fors (2018)	X				X	X	X		X
7	Liu, Wang and Koehler (2019)	X				X	X	X		X
8	Setiawan, Hamra, Jabu and Susilo (2018)			X	X	X				X
9	Aşık, İnce and Vural (2018)	X		X		X				X
10	Getu and Teka (2018)			X				X		X
11	Tseng (2016)	X				X				
12	Tseng (2018)	X		X		X	X			X
13	Tseng, Cheng and Lin (2011)				X	X	X	X		X
14	Tseng, Cheng and Yeh (2019)			X	X	X				X
15	Turgut (2017b)	X			X	X	X	X		X
16	Turgut (2017a)	X		X		X	X	X		X
TPACK studies that included other subject discipline areas such as social sciences education, science education etc.										
17	Baran and Uygun (2016)	X		X	X	X	X			
18	Baran, Canbazoglu, Albayrak and Tondeur (2019)	X				X	X			X
19	Chai, Koh and Tsai (2010)	X				X				
20	Dong, Sang, Chai, Koh and Tsai (2015)	X				X	X	X	X	X
21	Joo, Park and Lim (2018)	X				X		X		
22	Khine, Ali and Afari (2017)	X				X		X		
23	Redmond and Peled (2019)	X				X	X	X		X
24	Sang, Tondeur, Chai and Dong (2016)	X				X		X		

Lin, K.-C., Lee, I.-C., Cheng, C.-F., & Hung, H.-C. (2020). The Effects of Adopting Tablets and Facebook for Learning Badminton Skills: A Portfolio-Based WISER Model in Physical Education. *Educational Technology & Society*, *23* (4), 89–105.

The Effects of Adopting Tablets and Facebook for Learning Badminton Skills: A Portfolio-Based WISER Model in Physical Education

Kuo-Chin Lin[1], I-Chen Lee[2], Chih-Fu Cheng[3] and Hui-Chun Hung[4*]
[1]Center for Physical and Health Education, Si Wan College, National Sun Yat-sen University, Kaohsiung, Taiwan // [2]Department of Statistics, National Cheng Kung University, Tainan, Taiwan // [3]Department of Physical Education, National Taiwan Normal University, Taipei, Taiwan // [4]Graduate Institute of Network Learning Technology, National Central University, Taoyuan, Taiwan // linkuochin@mail.nsysu.edu.tw // iclee@mail.ncku.edu.tw // jeff@ntnu.edu.tw // hch@cl.ncu.edu.tw
[*]Corresponding author

(Submitted February 26, 2020; Revised April 24, 2020; Accepted September 21, 2020)

ABSTRACT: To improve the badminton skills learning of students, this study proposed the portfolio-based WISER model, which combines tablets for instantaneous recording and Facebook for e-portfolios. The 97 learners in the experimental group (EG) were taught with tablets and Facebook, and the 102 learners in the control group (CG) were taught with traditional teaching methods. The paired sample t test and ANCOVA were used for statistical analysis. The posttests of smashing and footwork in the EG and CG were both significantly higher than those on the pretests. The posttest scores for smashing were significantly higher in the EG than those in the CG. For pretest scores were higher than 7.66 or lower than 5.09, the posttest footwork scores were significantly higher in the EG than in the CG. The proposed portfolio-based WISER model can help students reinforce their understanding of badminton skills and improve their skill learning.

Keywords: Tablets, Social media, Badminton, Motor skill learning, e-portfolio

1. Introduction

Physical education (PE) is an essential subject for students' growth and health. Instructor meet the difficult challenge of balancing the complexity (typically one physical skill contains many detailed movements), spatiality (coordination of different body parts), precision, and speed of teaching contents (McKenzie & Lounsbery, 2013). The combination of technology and education is a current trend in PE. With the development of new technology, researchers now can apply 3D animation, multimedia websites, and online learning systems to assist PE teachers (Lan, Wang, & Chen, 2010; Papastergiou & Gerodimos, 2013; Zou, Liu, & Yang, 2012). The multimedia tools used in the aforementioned studies, however, must be operated in front of computers rather than during lectures. To solve this problem, our previous study applied tablets as instantaneous feedback tools and showed that tablets enabled the student to quickly comprehend how to correctly perform the skills being taught (Hung, Young, & Lin, 2018). Moreover, the related research has confirmed our previous study and highlighted the potential relevance of using video feedback in fostering motor learning engagement and motivation (Potdevin et al., 2018). Therefore, further research is needed to provide greater freedom to obtain visual feedback during the learning process.

In the meantime, social media has changed not only the means through which people communicate but also personal habits of sending, receiving, and sharing information (Magogwe, Ntereke, & Phetlhe, 2014; Wise, Skues, & Williams, 2011). According to a Statista report, more than 3.6 billion people worldwide use social media (Statista, 2020). Social media platforms, such as Facebook, have become daily essential tools for the younger generation growing up with the Internet and these platforms play an important role in many people's lives (Statista, 2020). For university students, Facebook helps maintain social relationships through continuous use (Raza, Qazi, & Umer, 2017).

Facebook has been explored in the application of education by many researchers. Some research has indicated that integrating Facebook into education has positive effects on student learning (Kabilan, Ahmad, & Abidan, 2010; Deng & Tavares, 2013; Magogwe et al., 2014). Magogwe et al. (2014) found that students often discuss course projects on Facebook, share course-related materials with their classmates, and ask questions to clarify lecture concepts. These results are also similar to those obtained by Deng and Tavares (2013), which demonstrated that Facebook helps to extend class discussions beyond the classroom. Kabilan et al. (2010) likewise indicated that using Facebook in class motivates students to learn, participate, and interact with their peers. Moreover, as a learning management system, Facebook is easily operated and accessible through mobile phones (Magogwe et al., 2014).

For this reason, this study adopts Facebook as a learning management system and proposes an innovative learning approach that uses Facebook Messenger to create e-portfolios of students' badminton skill learning. Through the interactive function of Facebook Messenger, instructors and learners can easily send one another information. At the same time, an e-portfolio is utilized for the collection of digital data regarding a student's learning progress, including texts, audio tracks, videos, and images. These digital data may also include feedback from experts or peers, the opinions of learners formed during the learning process, or later through introspection (Chang, 2014; Chu, 2009). Chang (2014) further pointed out that e-portfolios help students set learning objectives, accumulate knowledge, and consequently enhance their learning performance. In addition, Melograno (2000) claimed that student learning in PE needs to be evaluated in a more natural and performance-based manner, emphasizing the importance of recording students' growth in learning.

In view of the remarkable capability of tablets to provide instantaneous feedback, and the accessibility of Facebook Messenger, the aim of this research is to provide possible solutions to the existing problems in PE teaching, and eventually to facilitate badminton skill acquisition through the application of common handheld technologies and social media. This study intends to explore how tablets and Facebook can be integrated to form an innovative approach in an authentic learning setting. The main research objective is to understand the effects of the combination of tablets and Facebook to create e-portfolios on badminton learning. In order to reach this objective, an innovative approach is applied to PE courses to introduce a new perspective on PE teaching.

In this study, smashing and footwork were selected as the target learning skills due to their higher difficulty of acquisition. Smashing was also selected because it plays a key role in scoring in female doubles games (Tsai, Liao, & Chen, 2016). Footwork was chosen because the agility of footwork is the basic foundation of badminton (Chang, Chang, Chen, Ho, & Chen, 2016). Based on the research objective and the two badminton skills selected, this study poses two research questions:

(1) What are the effects of adopting tablets and Facebook on learners' badminton smashing performance?
(2) What are the effects of adopting tablets and Facebook on learners' badminton footwork?

2. Literature review

2.1. Motor skill acquisition in physical education

PE provides positive and profound effect for students in physical, lifestyle, affective, social, and cognitive domain to enhance their health (Bailey, 2006). PE focuses on learning through exploratory operations and requires repeated exercises of motor skills. Different from perception and cognitive skills, motor skills are a set of internal processes combined with repetitive exercises (Gagné, 1984; Newell, 1991; Schmidt & Lee, 1999). To gradually automate the skills, the acquisition of motor skills can be introduced from three stages: the cognitive stage, the association stage and the autonomy stage (Fitts & Possner, 1967).

In the cognitive stage, learners focus on executing skills and discovering what to do (Schmidt & Lee, 1999). Because this stage includes the learner's conveyance and acquisition of new information, this stage is also called the speech movement stage (Adams, 1971). In this stage, learners collect information through vision to learn and understand the purpose of the operation, and then gradually transfer the action into long-term memory. In this stage, learners' movements are not yet coordinated and stable, and they still do not know where their shortcomings are. Therefore, in this stage, using oral instructions and visual models can help learners recognize the main points of the action (Schmidt & Lee 1999). In PE class, the teacher's instruction, guidance, and execution of slow-motion drills are highly effective for students to cognitively understand the requirements and parameters of motor movement.

In the associative stage, the learner aims to make movement adjustments and string small movement skills together. After acquiring the basic concepts of movement in the cognitive stage, learners have understood the direction of improvement and continue to practice. This stage is also called the motor stage (Adams, 1971), because it aims to learn how to perform skills (Schmidt & Lee, 1999). After continuous practice, learners need to detect and correct errors in performance in order to perform smooth and stable movements. Therefore, learners need to transform their declarative knowledge into procedural knowledge, which means transforming what to do into how to do it. In PE class, the teacher's feedback should be more accurate and focus on the timing of the skills that the student is attempting to refine.

After continuous practice, the student can reach the autonomous stage. In this stage, the learners do not have to concentrate on skill and can perform a skill without paying attention to the action itself. They seldom consider the correctness of the action but react automatically, which means that athletic performance is largely automatic and the need for cognitive processing is small (Schmidt & Lee, 1999). Exercise during this period has entered long-term memory and players automatically adjust movement consistently and stably. There is still room for improvement, but the corresponding progress is slow. In PE class, students can still make small mistakes, and be able to find and correct them and produce their best performance.

PE curriculum only gives students a limited frequency and time for class. Therefore, the study indicates that teachers should promote physical exercise in their spare time. Pang (2010) pointed out that students' inability to obtain guidance outside of the course space and schedule are the drawbacks and challenges of traditional face-on-face PE in higher education. PE courses should clearly require students to explore opportunities for physical exercise before, during, and after school and within the community (McKenzie & Lounsbery, 2013).

2.2. Related research about using technology in physical education

Educational technology has been widely used in PE. From 2010 to 2012, educational technology, including online platforms, started draw significant research attention. Some researchers have used online learning as a platform for their studies. Huang, Chiu, Chin, Hsin, and Yu (2010) used an e-learning platform to provide sports materials and to demonstrate physical motions and exercises. The findings of their study indicated that multimedia content not only increase the efficiency of learning but also its appeal to students. Similarly, Papastergiou and Gerodimos (2013) used a web-based multimedia course to help undergraduate students acquire new knowledge about basketball. This study demonstrated that the combination of web-based multimedia courses and traditional face-to-face teaching is more effective than face-to-face teaching alone. Additionally, Zou et al. (2012) designed a Moodle course (an online learning management system) that provided videos of the world table tennis champions performing various strokes to help children learn how to play table tennis. The study demonstrated that students who participated in the Moodle course exhibited more knowledge and competence.

Since 2012, more technology products have been developed and used for educational research purposes and related studies have displayed greater research efficiency. Huang, Kuo, and Lin (2015) developed a golf-analysis system with a Wii balance board and Kinect to help detect golf-swing errors. Their study provided novice golf players with a reasonably priced and efficient golf-swing analysis system. Chen et al. (2013) also used Kinect, in combination with a computer-assisted self-training system, for yoga exercises, and developed a system capable of assisting a yoga practitioner in self-training, the performance of correct postures, and injury prevention. Aside from studies that have used the Kinect, Chen, Li, Lin, and Chang (2015) used a computer-aided system to train and improve the judgment skills of volleyball referees. Their findings showed that with the aid of their system the judgment performance of referees was improved. Yang, Li, and Xu (2017) also used a computer-aided artificial-intelligence-instruction system for basketball training in PE. The system in their study showed numerous advantages over traditional teachings, such as the liberation of human resources, reduced cost, energy conservation, and the modernization of education. Additionally, Chang, Ho, Wang, and Lin (2014) used a Wireless Inertial Measuring Unit (IMU) to measure body rotation in golf swings. The findings revealed that a wireless IMU is a feasible method for measuring body rotations. Lee et al. (2015) also used an IMU device for assessing lower limb locomotion, which demonstrated that IMU sensors exhibit potential capabilities for physical activities with different levels of intensity, such as walking, running, and jumping. Furthermore, Kitagawa and Ogihara (2016) used an IMU to develop a system for monitoring human gait. They did so by first identifying characteristics of gait, then by using a sensor attached to the dorsal of the feet of elderly individuals to measure foot trajectories. Yu, Wu, Wang, Chen and Lin (2020) also used tablets and motion-tracking sensors to help students to record their shot data and utilized the Moodle system as a reflective platform for students to review the practicing videos and rethink their performance.

The aforementioned studies in teaching and training all took time and required financial resources in the development of a new system. Therefore, some researchers have turned tablets, comparatively economical teaching tools, in their studies. Hung and Chen (2016) investigated the teaching of table tennis in PE by using tablets and discovered that combining tablets with traditional learning methods had beneficial effects on students' skill development. Moreover, our previous study (Hung et al., 2018) explored the ways in which tablets could be integrated into a badminton course to enhance the motivation and skills of learners. Furthermore, Hung et al. (2018) also indicated that the proposed WISER model – Watching, Imitating, Self-examining, Enhancing and Repeating – could be verified as an ideal pedagogical model for PE with mobile devices. In WISER model, students would watch the demonstration videos and imitate the skills. When students started to practice, the practicing process would be taken by tablets meanwhile. After practicing, students could improve themselves through reviewing the videos. They could raise the questions and discuss with the instructor about how to improve the skills. The instructor could also give proper advice to students, which could help students learn the skills effectively. Then, the students could practice and improve their skills repeatedly. The students who adopted the WISER model showed better performance in both motor-skill learning and motivation than those who were given the face-to-face approach. However, further pedagogical research is needed to examine the ways students can use video feedback technology for greater self-regulation (Hung et al., 2018). Studies indicate that student learning needs to be evaluated in a more naturalistic and authentic assessment manner (Melograno, 2000; Tolgfors, 2018). Therefore, this study searches for solutions to evaluate students' learning growth. Compared to the conventional assessment tools, a portfolio is a great tool for students to collect their learning outcomes in more authentic situations and to enhance their personalized PE experience. Therefore, this study aims to employ a friendly and familiar tool as a digital portfolio with tablets to enhance students' personalized PE learning experience and to prompt or reinforce student engagement in physical activity, sport, or fitness beyond the PE lesson.

2.3. Social media networks and education

Learning Management Systems have been widely used as a method for enabling students to engage with course learning materials. Innovative learning models are required to meet the social needs of a new generation of students. For these students, learning environments require connections to rich social environments that cannot be attained only through conventional learning management systems (McLoughlin & Lee, 2007). The rapid development of social media networks has created opportunities for innovation by allowing people to present themselves, share and receive information, and connect with one another (Manca & Ranieri, 2013; Al-Mashaqbeh, 2015).

Social media networks were initially designed for the purpose of social interaction between those who shared a campus or special-interest group. As informational and computational media, social media networks can enhance interactions between teachers and students outside of class. Currently, an increasing amount of research is focused on examining the potential of these networks in education and various pedagogical processes (Aydin, 2012; Ellison, Steinfield, & Lampe, 2007; Al-Mashaqbeh, 2015; Greenhow & Askari, 2017; Manca & Ranieri, 2016). Published literature reviews have focused mainly on the perceptions and experiences of college students (Aydin 2012; Manca & Ranieri, 2013) and higher education faculty (Forkosh-Baruch & Hershovitz 2012). Several studies have analysed data from student surveys to assess the benefits of Facebook as an educational tool. Social media tools can support student education by functioning as pedagogical tools for connectivity, social support, and information aggregation (Mazman & Usluel, 2010). Ellison et al. (2007) found that students' online social activities with Facebook help them to engage in their communities actively, especially for students with low self-esteem and low life satisfaction. Roblyer, McDaniel, Webb, Herman, and Witty (2010) reported that students were likely to be open to the potential of Facebook as a method for educational communication and mentoring. Moreover, applying the concepts of social software and Web 2.0 applications, Mcloughlin and Lee (2007) proposed 4 benefits of using social software tools as an educational method: (1) connectivity and social rapport; (2) collaborative information discovery and sharing; (3) content creation; and (4) knowledge and information aggregation and content modification. Moreover, some studies have indicated that using Facebook as the course management system is a cost-efficient and effective way to increase student involvement in discussions and out-of-class communication among instructors and students (Albayrak & Yildirim, 2015; Miron & Ravid, 2015; Raza et al., 2017). Therefore, Facebook is a social networking tool that allows groups and individuals to engage in peer-to-peer conversations.

In tandem with the aforementioned research, this study aims to combine the use of tablets and Facebook as a friendly portfolio tool in PE to assess their effects on badminton learning. Using Facebook and tablets enables this study to lower the cost, appeal to users, and save more time because no time was spent on system development.

3. Methods

3.1. Portfolio-based WISER model

The features of tablets and Facebook allowed this study to enrich the content of lectures. A Facebook Group was set up for learners to acquire the learning materials during and after class. In the meantime, to construct the learners' e-portfolios, the instructor recorded each learner's motions during class and uploaded them to their individual Facebook account through Messenger after class. Therefore, this study improved upon our previous WISER model (Hung et al., 2018) by developing the "portfolio-based WISER model," as illustrated in Figure 1. The original five-step WISER model was enhanced by portfolios to achieve a more flexible and personalized pace of PE learning. Students could obtain self-reflection from their video-based portfolios and the interaction with the instructor. Therefore, they could monitor their growth in skill learning. Before conducting the portfolio-based WISER model, the instructor illustrated the whole process in detail for students to make them clear about the process. Also, the instructor assisted students in following the model in a correct way so they could practice.

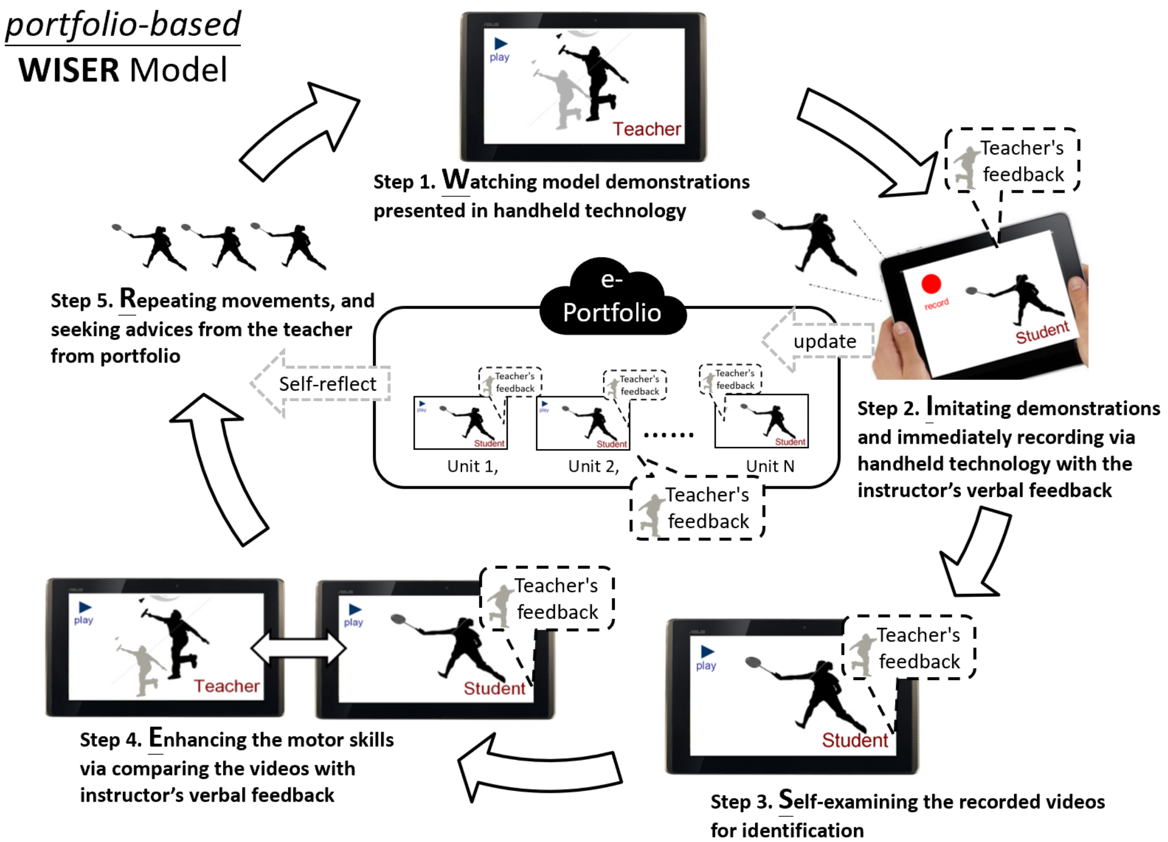


Figure 1. Portfolio-based WISER model for PE learning

Step 1. Watching model demonstrations presented with handheld technology: The students needed to watch the digital videos of different basic skills, which were demonstrated by the teacher via tablets. The students could go forwards and backwards, and replay the videos.

Step 2. Imitating demonstrations and immediately recording via handheld technology with the instructor's verbal feedback: The students needed to imitate the same movements according to the videos they just watched. At the same time, the students were divided into groups, and their practice processes were recorded by the instructor along with the instructor's verbal feedback via the recording function of the tablets. The instructor gave corresponding feedback based on the different performances of each student following the skill grading criterion (see Table 1).

Step 3. Self-examining the recorded videos for identification: After recording the practice videos on the tablets, the students could watch them instantly on the screen of their tablets. These videos enabled students to identify the flaws of their own movements according to the teacher's verbal feedback.

Step 4. Enhancing motor skills by comparing the videos with the instructor's verbal feedback: The students needed to compare the practice videos and teacher's demonstration videos to further reflect on the differences between them so that they could follow and refer to the teacher's verbal feedback in the video as a supplementary reference.

Step 5. Repeating movements and seeking advice from the teacher through portfolio: After understanding the differences between the practice videos and the teacher's demonstration, the students could practice their movements repeatedly and then seek further advice from the teacher face-to-face. In the portfolio-based WISER model, the teacher can give comprehensive feedback to students by referring to the students' portfolios from different weeks.

3.2. Participants

This study divided the research participants into the experimental group (with the portfolio-based WISER model) and the control group (with traditional teaching methods). The lecturing content of both groups was identical. The number of the participants was 199 in total, who were students from six badminton classes at the university where the researcher is affiliated. The researcher assigned three classes as the control group and three classes as the experimental group before the experiment started. Each class was grouped into four groups on a random basis, and each group included 8 to 9 students. Participants were grouped according to the university's course information system; three classes formed the control group in one semester (Class 1 = 30; Class 2 = 36; Class 3 = 36; *N* = 102), and the other three classes formed the experimental group in the next semester (Class 1 = 29; Class 2 = 33; Class 3 = 35; *N* = 97), for a total experimental period of one year. During the experiment, some learners failed to complete the smashing or footwork tests, and some withdrew from the class.

3.3. Experimental procedures

The experiment was based on a quasi-experimental non-equivalent pretest/posttest control/comparison group design. In the beginning, this study uploaded pre-recorded teaching materials of smashing and footwork skills demonstrated by professional Taiwanese badminton players onto a tablet and Facebook. Next, students in the control group were taught using traditional teaching methods and students in the experimental group were taught using tablets and Facebook. Before and after this step, participants took a pretest and a posttest. The timeframe of this experiment consisted of two semesters, each including one week of pretesting, eight weeks of lectures, and one week of posttesting. The same instructor led both experiment group and control group in both semesters. Figure 2 shows the experimental procedure.

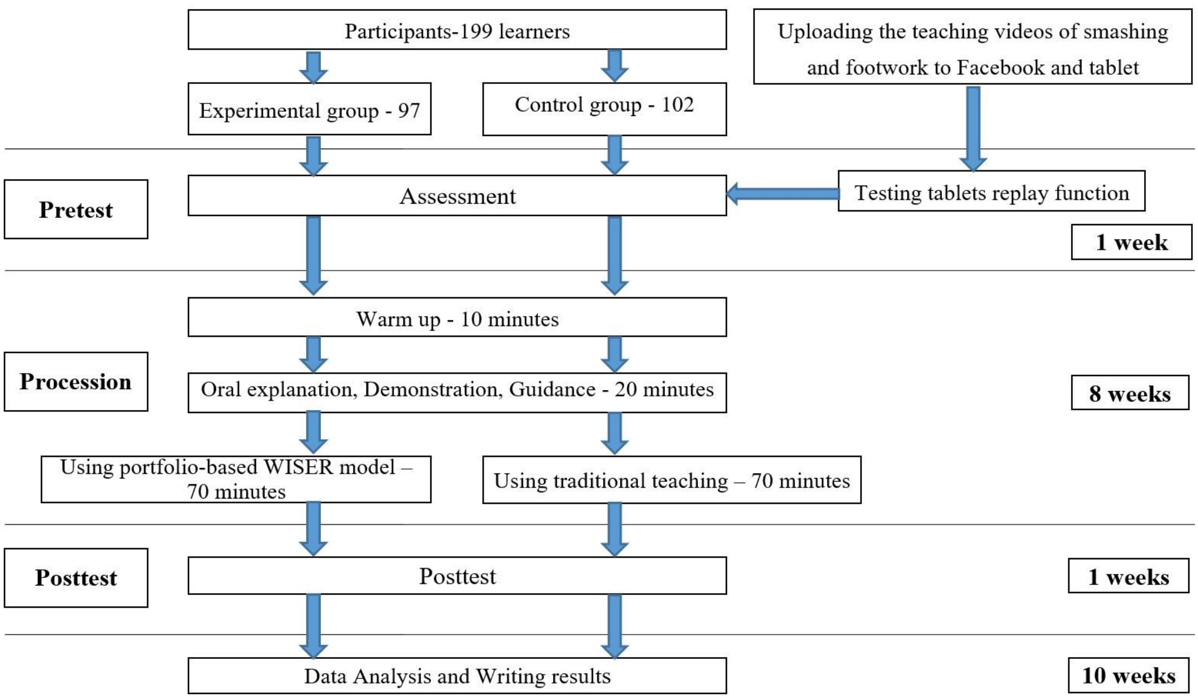


Figure 2. Experimental procedure

3.3.1. Procession step: Control group

The control group was given a 100-minute lesson for eight consecutive weeks during the semester. These lessons included a 10-minute warm-up, 20-minute lecture, and 70-minute group practice. During the 20-minute lecture, the instructor verbally explained and physically demonstrated smashing and footwork skills. For the next 70 minutes, the instructor provided each learner with one-on-one verbal and demonstrative feedback and guidance until the class was finished. The instructor rotated to each group to observe every learner's learning process and offered the comprehensive feedback through verbal and physical demonstration one on one.

3.3.2. Procession step: Experimental group

The experimental group was also given 100-minute lessons for eight consecutive weeks during the semester; these lessons included a 10-minute warm-up, 20-minute lecture, and 70-minute group practice. The lecture was executed using tablets and Facebook content that had been uploaded at the beginning of the semester. The primary purpose for the teaching strategy was to provide students with a model demonstration videos that could be used to compare with their personal practice videos on the Facebook Group.

During the 70-minute group practice, the teacher adopted the portfolio-based WISER model to instruct students. Each group containing 8 to 9 students was assigned to one tablet. The teacher rotated to each group to provide the comprehensive feedback regarding the movements during each student's practice process, and both comprehensive feedback of the teacher and the practice process of the student were video-recorded simultaneously as a practice video. After the teacher finished providing the feedback, the students could watch the practice video to receive the comprehensive feedback for the second time with the replay function of tablets, or raise some problems from videos to the instructor personally.

After the seventy-minute practice time, the practice videos were uploaded to each student's Messenger account. The students could watch the practice video for after-class review and compare it with the model demonstration videos on the Facebook Group (as shown in Figure 3). This process enabled students to receive comprehensive feedback for the third time. Also, should the students have any concerns regarding the movements, they could discuss with the teacher with Messenger, which could enhance the interaction between the teacher and students. The e-portfolio was created for students to record their entire learning progress on their personal Messenger account (as shown in Figure 4).

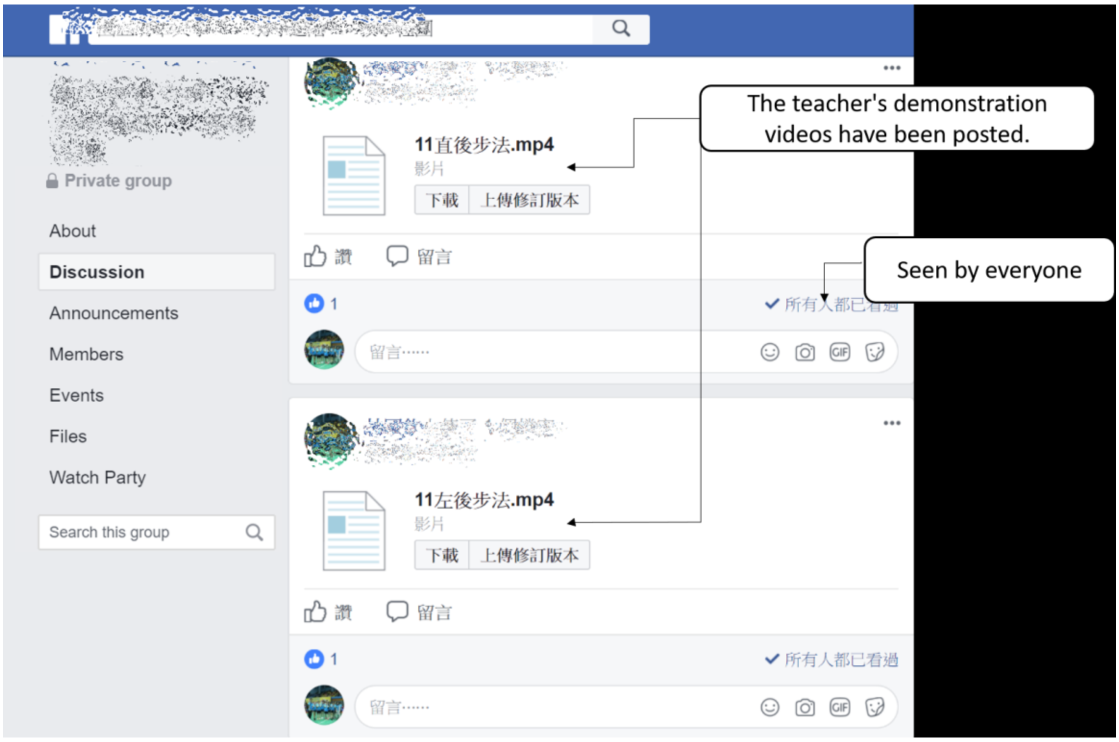


Figure 3. Facebook Group for badminton skill learning

Throughout the eight-week experiment, participants in the experimental group could use tablets to make movement adjustments via real-time visual feedback and use the e-portfolio for after-class-reviewing via Facebook Messenger. Figure 3 and Figure 4 present the Facebook teaching group and the student's e-portfolio, respectively. This study applies Facebook Group and Facebook Messenger for students' badminton learning. In the Facebook Group, the teacher uploaded the model demonstration videos. Therefore, students could also review the teacher's model demonstration before and after class. In the Facebook Messenger, the student received his/her weekly practice videos after class (Figure 4). Therefore, students have video-based portfolio made with their weekly videos. Moreover, after students watched their practice video, students can seek further personal guidance from the teacher according to the practice video. In the portfolio-based WISER model proposed in this study, we apply Facebook to facilitate the relationship not only from instructor to learners but also from learners to instructor. Therefore, beyond one-way teaching, this study uses Facebook as a social networking tool from two-way teaching.

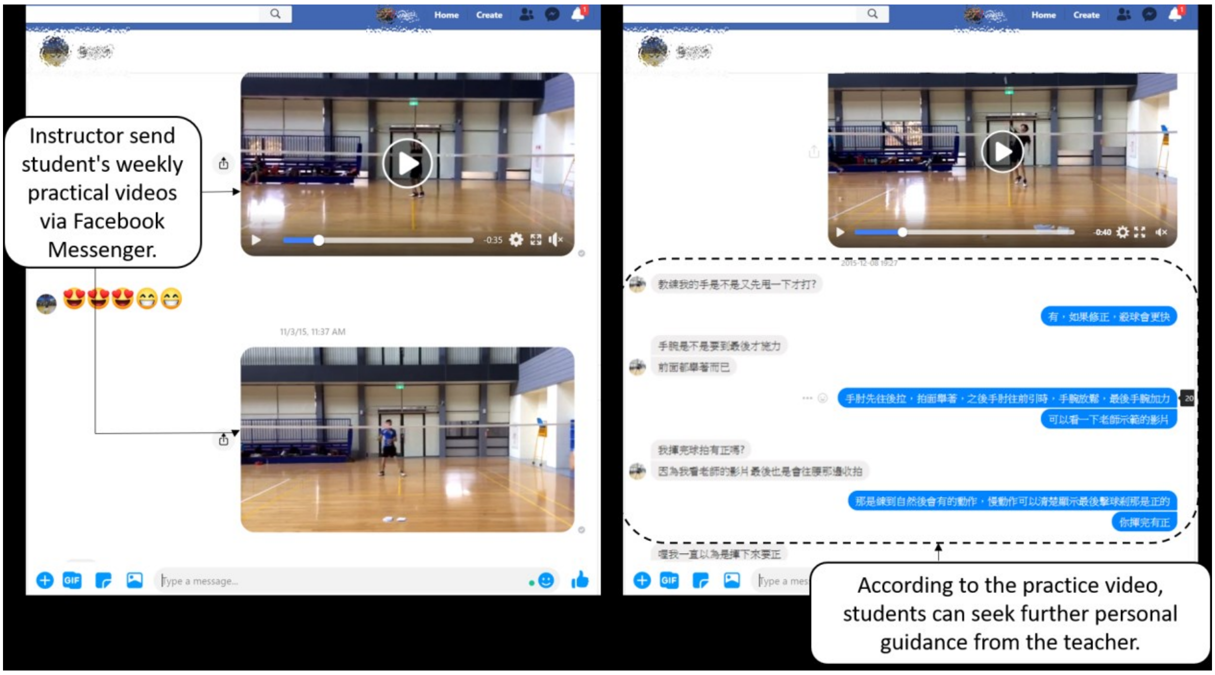


Figure 4. Learner's e-portfolio

3.4. Instruments

Three research instruments were used in this study. The first was a smashing learning test adapted from badminton skill test (Lin, Hung, & Young, 2014), the "high clear learning test grading criteria"; this included seven criterions for grading, as displayed in Table 1. The second instrument was a footwork-learning test with nine criterions. The criteria used in the smashing and footwork-learning test (as shown in Table 1) were determined by three professors certified as A-rank players and national coaches by the Chinese Taipei Badminton Association (CTBA). Therefore, these two tests have high content validity. When a learner's one movement matched one of the criterions, he or she would get one score for the test.

Table 1. Grading criterions of the smashing and footwork-learning test

Skill	Smashing	Footwork
Grading criterions	• The elevation of both elbows • The balance of both arms • 90 degree turn to the right/left side • The rotation of the shoulders • The snap of the wrist • The follow-through of the shoulders • The fluency of movements	• Forehand backcourt front cross footwork • Backhand backcourt back cross footwork • Backcourt straight back cross footwork • Forward forehand back cross footwork • Forward backhand front cross footwork • Frontcourt straight front cross footwork • Forehand side footwork • Backhand side footwork • The fluency of movements

To determine reliability, the researcher and one CTBA player assigned grades according to the smashing and footwork criteria. Kendall's coefficient of concordance was also used to determine the reliability of the two tests. The pretest and posttest scores for smashing and footwork obtained using Kendall's coefficient of concordance were .89, .84, .93, and .93, respectively.

3.5. Data analysis

Three data sets were used for the data analysis badminton skills learning. To make sure that the students were assigned to each group without any preference, an independent sample *t* test was utilized to gauge if there was a significant difference between the pretests of both groups. Second, in order to understand different progress levels of both skills in each group, a paired sample *t* test was deployed to calculate whether the posttests were significantly better than the pretests. Third, if the posttests were significantly better than the pretests, an analysis of covariance (ANCOVA) was be conducted to determine whether a significant difference existed between the experimental group and control group. A homogeneity test was also conducted (where the scores of the pretest were the covariate and the scores of the posttests were the dependent variable) to determine similarities between the two groups. The significance level for all statistical analysis was $\alpha = .05$.

4. Results

An independent *t* test was first conducted in the present study for the purpose of knowing whether learners were assigned without any preference for each group. The result was that there was no significant difference in the pretest scores between the two groups (*p*-value = .13 for smashing and *p*-value = .33 for footwork). The learning outcomes of smashing and footwork in each group are be discussed individually below.

4.1. Badminton skills learning outcomes of EG and CG

Table 2 summarizes the results of the paired sample *t* test of the pretest and posttests scores in the experimental group and control group. It was found there was a significant difference between the pretest and posttests scores in both EG and CG, and the posttests scores were higher than the pretest scores in both groups. This means that the learners improved their skills after they were taught with either teaching method for badminton skills.

Table 2. Summary of the paired sample t test of the pretest and posttest scores in each group

Testing skills	Group	Testing time	Mean	*SD*	*n*	*t*	*p*
Smashing	Control	Pretest	3.76	.99	99	14.41	$< .000^{***}$
		Posttest	5.27	.96	99		
	Experimental	Pretest	3.49	1.41	97	16.24	$< .000^{***}$
		Posttest	5.43	1.02	97		
Footwork	Control	Pretest	2.51	2.19	102	14.84	$< .000^{***}$
		Posttest	5.64	1.88	102		
	Experimental	Pretest	2.83	2.38	95	21.40	$< .000^{***}$
		Posttest	7.80	.70	95		

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

4.2. Smashing

Before conducting an ANCOVA, we first tested for any difference in pretest scores between the experimental and control groups. For the smashing test, the results of the test for homogeneity in smashing skills demonstrated that the interaction between covariates (pretest scores) and dependent variables (posttest scores) had no significance, with the *p*-value = .96. This indicated that no significant difference existed on the slopes of regression of the pretest against posttest scores between the two groups; hence, ANCOVA could be conducted.

After conducting the homogeneity test for each group to exclude the effects of pretest scores on posttest results, ANCOVA analysis could be carried out. The results indicated that there was a significant difference between the posttest scores of the experimental group and the control group, with *p*-value = .032 (Table 3). Additionally, average scores after adjustments were 5.48 for the teaching method involving tablets and Facebook and 5.22 for the traditional teaching method. These results indicated that using the teaching method involving tablets and Facebook led to more significant improvements in smashing skills than using the traditional teaching method. In Figure 5, the blue and red lines are the fitted lines of the relationship between pretest and posttest scores for the experimental and control groups, respectively. It was found that both fitted lines are almost parallel, where the homogeneity of the slopes for both groups existed. In addition, the average posttest scores of the experimental group were higher than those of the control group.

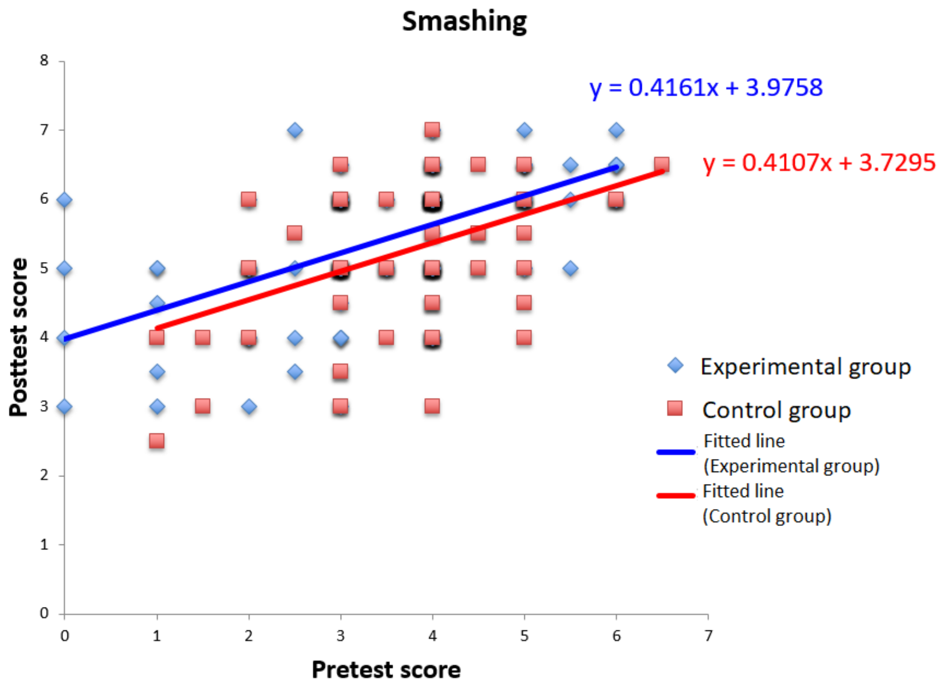


Figure 5. The relationship between the pretest and posttest scores of the two teaching groups

Table 3. ANCOVA table of smashing tests

Source	Sum of squares	Degrees of freedom	Mean square	*F*	*Sig.*	Partial Eta squared
Pretest scores (smashing)	49.29	1	49.29	67.31	.000	.259
Group	3.43	1	3.43	4.68	.032*	.024
Error	103.83	106	0.98			

Note. *p <.05.

4.3. Footwork

For the footwork test, the results of the homogeneity test for footwork demonstrated that the interaction between covariates (pretest scores) and dependent variables (posttest scores) was significant, with $p < .001$, indicating a difference in the slopes of the pretest scores against the postest scores between the experimental and control groups. From Figure 6, the non-homogeneity was shown by the blue and red lines which were the fitted lines of the relationship between pretest and posttest scores for the experimental and control groups, respectively. Unlike Figure 5, the two fitted lines are not parallel so that the assumption of homogeneity of the slopes for both groups failed when conducting an ANCOVA. Because the assumption of homogeneity was not satisfied, ANCOVA could not be conducted. Instead, the Johnson-Neyman method was used to adjust the results. The Johnson-Neyman technique provides additional information by establishing regions of insignificance for group difference at a specific point on the covariate (Johnson & Neyman, 1936; D'Alonzo, 2004).

According to the procedure of D'Alonzo (2004) and Johnson (2016), the insignificance difference of two groups was found at the pretest scores occurring in the range of 5.09 and 7.66. That is, when the pretest score is higher than 7.66 and lower than 5.09, a significant difference exists between the two groups in the posttest. In Figure 6, when learners' pretest scores were higher than 7.66, the experimental group's posttest scores were significantly higher than those of the control group. When learners' pretest scores were lower than 5.09, the experimental group's posttest scores were significantly higher than those of the control group. When learners' pretest scores were between 5.09 and 7.66, no significant difference was evident between the posttest scores of the two groups.

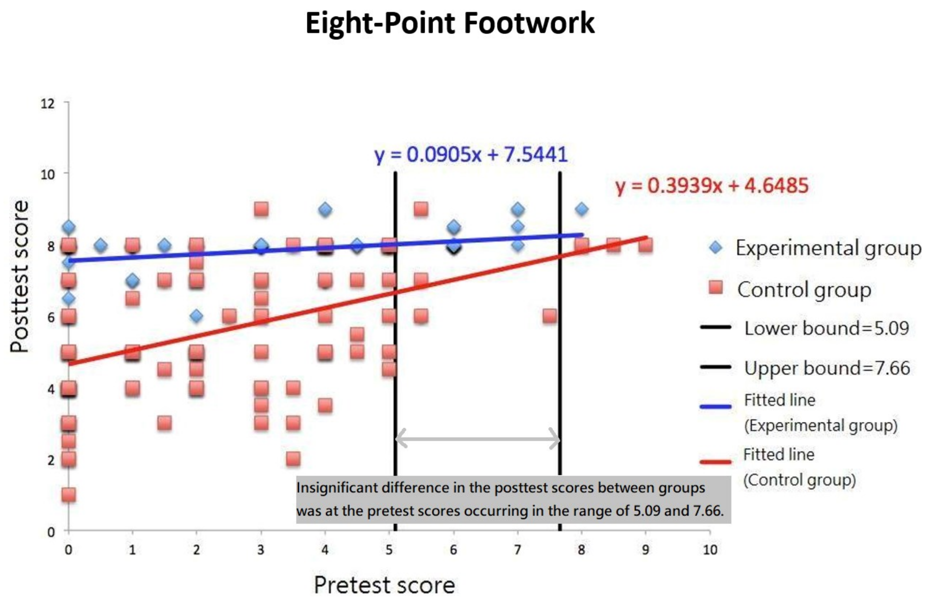


Figure 6. The results after employing the Johnson-Neyman method

5. Discussion

In traditional teaching without any recording devices, the learner is unable to see his/her mistakes because there is no recording of their performance available during class, and they must therefore rely on their memory alone outside of class. In this study, with the use of tablets, the performance of the student and the instructor's instant feedback were recorded as a performance video. This teaching method can be differentiated from past studies using tablets that recorded only the performance of students (Hung & Chen, 2016; Hung et al., 2018; Yu et al., 2020).

Furthermore, to create an e-portfolio, the present study uploaded the performance videos onto Facebook Messenger in each student's account. A student could review his/her performance videos regularly after class. If students had any confusion about the smashing or footwork skills, they could discuss questions with the instructor on Facebook Messenger through video or text (as shown in Figure 7). This teaching method increased the interaction, and helped students keep track of their learning progress each week and avoid making the same mistakes. The proposed method makes instruction more diverse and flexible. Students learn the main points of skills efficiently and deeply in advance. Also, e-portfolios are highly portable and easy operation. This study contributed empirical evidence to support past research utilizing Facebook as a course management system to increase student communication and involvement between the instructor and students in and out of class (Albayrak & Yildirim, 2015; Raza et al., 2017). In particular, students who have low self-esteem and satisfaction with life are more likely to communicate with the instructor after PE class (Ellison et al., 2007). Moreover, the results demonstrated that social networks can be an incredibly useful instructional tools because instructors and learners can share information and engage in discussion with one another through Messenger. Teaching methods that utilise social networks, when compared with traditional teaching methods, result in greater connectivity between instructors and learners (Al-Mashaqbeh, 2015; Manca & Ranieri, 2013; Mazman & Usluel, 2010).

For both the EG or CG, for smashing or footwork, with or without the portfolio-based WISER model, the posttest scores were all significantly better than the pretest scores, indicating students in the CG could still progress with only the adoption of the traditional teaching method. Additionally, with the aid of the portfolio-based WISER model, the smashing posttest scores of the EG were all significantly better than those of CG. However, the footwork posttests of the EG were significantly better than those of the CG only when the pretest scores were lower than 5.09 or higher than 7.66.

Furthermore, the present study explored the key points of smashing and footwork learning and investigated how the portfolio-based WISER model contributed to helping students with learning of both of these badminton skills. Smashing learning requires the understanding of the fluency and coordination of the shoulder, elbow, and wrist. In particular, the students need to know how to correctly use the strength from the shoulder, elbow, and wrist in sequential order when performing smashing to avoid injuries. On the other hand, compared to smashing learning, footwork learning only relies on the explicit instructions related to stepping with the left or right foot and jumping or not jumping in sequential order to alternate positions.

Due to the complexity of movements, this study designed a portfolio-based WISER model for use during and after classes. This model helped students learned the skills. During the 100-minute classes, students in the EG practiced the skills and the instructor provided instant feedback at the same time. In addition, students could review the practicing videos a second time after practicing. Then, the instructor and students could discuss the problems on Messenger to have more interaction in the instruction. In addition, based on the portfolio-based WISER model, the after-class reflection is also another important strategy to make promote skill learning. The idea is similar to developing an e-portfolio learning environment to support pre-service teachers' reflection of learning (Roberts, Maor & Herrington, 2016). As shown in Figure 7, the first and the second student confirmed their learning conditions through their extra practice videos.

The first student took a video of herself practicing and sent it to the instructor through Messenger to discuss the problems she saw. She mentioned that *"I could still see that the elbow was in the wrong position which was too low, and then I raised the elbow back up before hitting the shuttle."* Furthermore, she also raised two questions about the movements. *"Did I stretch my shoulder so much in the backswing that all the sub-motions were influenced?" "Is it possible to lose extra power or cause a sports injury when stretching the shoulder too much in the backswing?"* This example shows us the portfolio-based WISER model can supplement the after-class interaction, which could not be achieved through the WISER model in the past. The interactive approach was vivid and helpful, allowing the instructor to check if the movements were correct or not immediately and tell students whether or not incorrect movements could cause injuries to the body in advance.

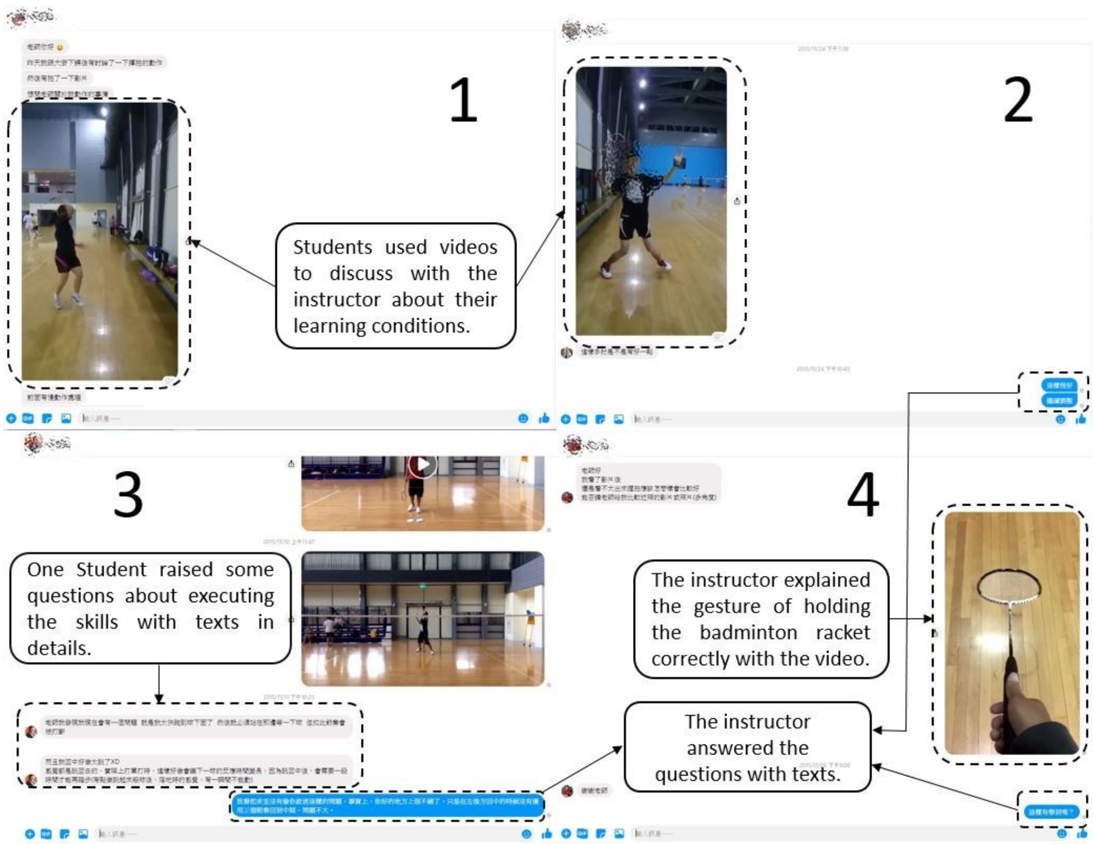


Figure 7. Interaction between the instructor and students on Messenger

From the weekly portfolio videos on Messenger, the instructor noticed the second student's (Figure 7) body and arms were overstretched in class. After watching the practice video, the student reflected on the instruction in detail. The student recorded an improved practice video and sent it to the instructor through Messenger. He asked the instructor, "*Did the swing of elbow improve?*" to confirm the correction of movements in advance. The instructor also provided positive feedback. In another example, the third student raised questions about executing the skills via texts. The student understood the importance of footwork and asked the instructor some questions after watching the videos on Messenger after class. He mentioned that "*I thought that I ran so fast that I reached below the shuttle too early. I had to stand there to wait for the shuttle, and the tempo was interrupted.*" He also said that, "*I found I jumped so much that I almost jumped back to the court centre and the hitting tempo felt wrong.*"

The above questions show that students raised these questions after reviewing the videos, which allowed students to reflect on how to apply the learning from class to practical game playing. In addition, the instructor also replied to students' questions with videos to make the explanations more vivid and concrete. The results from the experiment show that the proposed method made instruction diverse, flexible and interactive. Students learned the main points of skills efficiently and deeply in advance.

To further discuss how this study applied the portfolio-based WISER model in PE learning, Figure 8 shows the pedagogies and learning process. In the traditional class-based PE learning process, the pedagogy is an instructor-centered approach. In the portfolio-based WISER PE learning process, which is a learner-centered approach, convenient e-portfolios were established to provide more chances for students to reflect on the learning process and enhance the frequency of interaction between the students and instructor. This kind of pedagogy brings new chances and contributions to PE teaching.

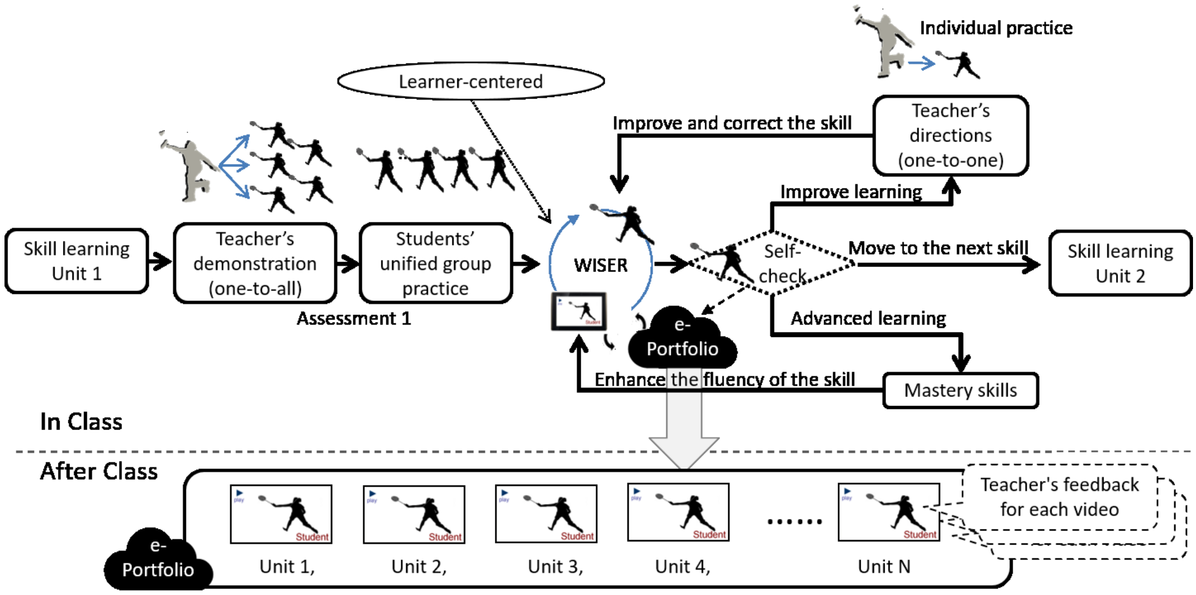


Figure 8. The pedagogies of portfolio-based WISER model in PE class

Fitts and Posner's (1967) stages of learning on motor skills acquisition have a long history in traditional PE. However, research (Davids, Bennett, & Button, 2008) also addresses a challenge of traditional PE, in that students only have a limited amount of time to learn with the teacher. However, to achieve a high level of expertise, students need to spend a lot of time on conscious practice. At the same time, the instructor has to monitor the individual progress of the students at different levels over time. Therefore, to deal with these challenges, this study further discusses how the proposed pedagogies of the portfolio-based WISER model could be exploited in each step of the motor skills acquisition process (Table 4).

In the cognitive stage, students need to understand what to do, therefore, the teacher needs to perform the correct demonstration in front of students (Assessment 1 in Figure 8). In this stage, the development of skill begins with clear instructions for practice. However, it might be challenging for students to comprehend instructions in the limited time during class (Kee, 2019). Therefore, students need to follow Step 1 of the portfolio-based WISER model to watch model demonstration videos with tablets. In this step, students can acquire visual and verbal knowledge and pay attention to the sequence and how to perform the motor movement correctly. Students need to imitate the same actions based on the video and pay attention repetitively to the object of focus.

In the associative stage, students aim to translate declarative knowledge into procedural knowledge to perform the skill after students understand the instructions in the cognitive stage. Timely feedback gives important support during the learning process. Moreover, it is helpful to receive verbal feedback from the teacher to ensure that students are practicing according to instructions (Kee, 2019). Therefore, they need to follow steps 2 to 4 of the portfolio-based WISER model. They need to imitate the same movement according to the videos they just watched. After recording the practice videos on the tablets, students can watch them instantly on the screen of their tablets. Moreover, students need to compare the practice videos with the teacher's demonstration video to further reflect on the differences. Therefore, students practice their movements repeatedly according to the teacher's oral feedback to find out the defects of their movements. After that, students should be familiar with the basic instructions of the practice.

In the autonomous stage, students try to automate their performance. In this stage, it is difficult to make major breakthroughs based on student performance. Students have internalized the skills automatically. Some practical suggestions for guiding students' skills development according to the individual characteristics are needed (Kee, 2019). Therefore, students need to follow Step 5 of the portfolio-based WISER model. Students can practice the movements repeatedly, and they can also seek further individualized advice from the teacher through Messenger. In addition, in order to gradually improve and continue to see students' progress, students need to refer to the personal portfolios. After extensive practice, students are expected to perform this skill automatically without paying attention and clear instructions.

Table 4. The portfolio-based WISER model facilitates motor skill learning in the following motor acquisition stages

Stage	Purpose	The related pedagogy design
Cognitive (Verbal-motor stage)	What to do (Gathering information)	• Teacher's demonstration • Step 1. Watching model demonstrations presented with handheld technology
Associative (Motor stage)	From what to do to how to do (Putting actions together)	• Step 2. Imitating demonstrations and immediately recording practice via handheld technology with the instructor's verbal feedback • Step 3. Self-examining the recorded videos for identification • Step 4. Enhancing motor skills by comparing the videos with the instructor's verbal feedback
Autonomous (Automatic stage)	Think minimally and react automatically (Extensive practice)	• Step 5. Repeating movements, and seeking advice from the teacher from the portfolio • Referring to their personal portfolios and seeking individualized advice after class

6. Conclusions

This study utilized the portfolio-based WISER model to help students reinforce their understanding of badminton skills, and the results indicated that using technology as a teaching tool is beneficial to badminton learning in PE. The key benefits of employing the portfolio-based WISER model in PE courses are described as follows:

(1) Movement adjustment via instantaneous recording and feedback:

The recording function of tablets enabled the instructor to instantly record and simultaneously provide verbal feedback to the learners. With the recordings, students were able to constantly adjust and improve their own movements by reviewing the videos.

(2) Connectivity and information sharing:

This study used Facebook Messenger to create an e-portfolio for each learner. Because learners were already actively using Facebook and were therefore comfortable operating the technology, it was relatively easy to make badminton learning in a social media setting a part of their routine. With a closed group on Facebook, the instructor and learners could easily share learning materials such as videos, references, links, and resources with other group members. In addition, all group members, including learners and the instructor, could create, organize, and share content with one another.

In this study, we applied tablets and Facebook to PE teaching and found that tablets and Facebook can transform the relationship between the instructor and learners from one-way teaching to two-way learning with more flexibility. However, if numerous learners are present in one class, some may be left behind. Therefore, future research should consider developing an automatic feedback system to assist the instructor. This would increase the frequency of feedback and enhance learning efficiency.

Acknowledgement

This work was supported by the Ministry of Science and Technology of Taiwan under contract numbers, MOST108-2410-H-110-055- and MOST108-2511-H-008-017-. The authors would like to thank all the people who took part in this study.

References

Adams, J. A. (1971). A Closed - loop theory of motor learning. *Journal of Motor Behavior, 3*(2), 111-150.

Albayrak, D., &Yildirim, Z. (2015). Using social networking sites for teaching and learning: Students' involvement in and acceptance of Facebook® as a course management system. *Journal of Educational Computing Research, 52*(2), 155–179. doi:10.1177/0735633115571299

Al-Mashaqbeh, I. F. (2015). Facebook applications to promote academic engagement: Student's attitudes towards the use of Facebook as a learning tool. *International Journal of Modern Education and Computer Science, 7*(11), 60–66. doi:10.5815/ijmecs.2015.11.07

Aydin, S. (2012). A Review of research on Facebook as an educational environment. *Educational Technology Research and Development, 60*(6), 1093–1106. doi:10.1007/s11423-012-9260-7

Bailey, R. (2006). Physical education and sport in schools: A Review of benefits and outcomes. *Journal of School Health, 76*(8), 397–401. doi:10.1111/j.1746-1561.2006.00132.x

Chu, P. C. (2009). Reflections on the implementation of portfolio in library user education. *Bulletin of Library and Information Science, 68*, 66-79. doi:10.6575/JoLIS.2009.68.05

Chang, C. C. (2014). Effect of e-portfolio on students' knowledge amassing. *Curriculum & Instruction Quarterly, 17*(4), 207-227. doi:10.6384/CIQ.201410_17(4).0008

Chang, C. H., Ho, C. S., Wang, M. S., & Lin, G. C. (2014). Parameter analysis of body rotation during golf putting by using wireless wearable device. *Journal of Leisure, Tourism, Sport, & Health, 4*(2), 81-90.

Chang, C. Y., Chang, C. H., Chen, K. C., Ho, C. S., & Chen, Y. R. (2016). The Effect of footwork and response capability in elementary school badminton players by six weeks agility training. *Sports Coaching Science, 44*, 57-66. doi:10.6194/SCS.2016.44.07

Chen, C. F., Li, C. C., Lin, I. C., & Chang, C. C. (2015). The Effect of the computer-aided system for volleyball referees' executive judgment. *Sports Coaching Science, 40*, 75-87. doi:10.6194/SCS.2015.40.06

Chen, H. T., He, Y. Z., Chou, C. L., Lee, S. Y., Lin, B. S. P., & Yu, J. Y. (2013). Computer-assisted self-training system for sports exercise using Kinects. In *Proceeding of the IEEE International Conference on Multimedia and Expo Workshops (ICMEW)* (pp. 1-4). doi:10.1109/ICMEW.2013.6618307

D'Alonzo, K. T. (2004). The Johnson-Neyman procedure as an alternative to ANCOVA. *Western Journal of Nursing Research, 26*(7), 804–812. doi:10.1177/0193945904266733

Davids, K., Bennett, S., and Button, C. (2008). *Dynamics of skill acquisition: A constraints-led approach*. Champaign, IL: Human Kinetics.

Deng, L., & Tavares, N. J. (2013). From Moodle to Facebook: Exploring students' motivation and experiences in online communities. *Computers & Education, 68*, 167–176. doi:10.1016/j.compedu.2013.04.028

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The Benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication, 12*(4), 1143–1168. doi:10.1111/j.1083-6101.2007.00367.x

Fitts, P. M., & Posner, M. I. (1967). *Learning and skilled performance in human performance.* Belmont, CA: Brock-Cole.

Forkosh-Baruch, A., & Hershkovitz, A. (2012). A Case study of Israeli higher-education institutes sharing scholarly information with the community via social networks. *The Internet and Higher Education, 15*(1), 58–68. doi:10.1016/j.iheduc.2011.08.003

Gagné, R. M. (1984). Learning outcomes and their effects: Useful categories of human performance. *American Psychologist, 39*(4), 377–385. doi:10.1037/0003-066X.39.4.377

Greenhow, C., & Askari, E. (2017). Learning and teaching with social network sites: A Decade of research in K-12 related education. *Education and Information Technologies, 22*(2), 623–645. doi:10.1007/s10639-015-9446-9

Huang, C. H., Chiu, C. F., Chin, S. L., Hsin, L. H., & Yu, Y. P. (2010). A sports e-learning platform: Teaching and learning by using multimedia contents. In *Proceedings of the 3rd IEEE International Conference on Ubi-Media Computing* (pp. 222-226). doi:10.1109/UMEDIA.2010.5544460

Huang, S. Y., Kuo, K. P., & Lin, Y. H. (2015). A Golf swing analysis system using Wii balance board and Kinect sensors for novice players. *Multimedia Tools and Applications, 74*(23), 10679-10696. doi:10.1007/s11042-014-2198-5

Hung, H. C., Young, S. S. C., & Lin, K. C. (2018). Exploring the effects of integrating the iPad to improve students' motivation and badminton skills: A WISER model for physical education. *Technology, Pedagogy and Education, 27*(3), 265–278. doi:10.1080/1475939X.2017.1384756

Hung, S. W., & Chen, W. C. (2016). A Study of integrating metacognitive strategy and Pad into physical education on the learning effectiveness of table tennis skills. *Journal of Taiwan Sport Pedagogy, 11*(2), 55-79. doi:10.6580/JTSP.2016.11(2).04

Johnson, P. O., & Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoirs, 1*, 57–93.

Johnson, T. R. (2016). Violation of the homogeneity of regression slopes assumption in ANCOVA for two-group pre-post designs: Tutorial on a modified Johnson-Neyman procedure. *The Quantitative Methods for Psychology, 12*(3), 253-263. doi:10.20982/tqmp.12.3.p253

Kabilan, M. K., Ahmad, N., & Abidan, M. J. Z. (2010). Facebook: An Online for learning of English in institutions of higher education? *The Internet and Higher Education, 13(4)*, 179–187.

Kee, Y. H. (2019). Reflections on athletes' mindfulness skills development: Fitts and Posner's (1967) three stages of learning. *Journal of Sport Psychology in Action, 10*(4), 214–219. doi:10.1080/21520704.2018.1549640

Kitagawa, N., & Ogihara, N. (2016). Estimation of foot trajectory during human walking by a wearable inertial measurement unit mounted to the foot. *Gait Posture, 45*, 110-114. doi:10.1016/j.gaitpost.2016.01.014

Lan, H. C., Wang, H. C., & Chen W. C. (2010). Investigation of multimedia-assisted instruction on 8-movement Taijiquan's learning effects for the college students. *Physical Education Journal, 43*(3), 109-125. doi:10.6222/pej.4303.201009.1008

Lee, Y. S., Ho, C. S., Shih, Y., Chang, S. Y., Robert, F. J., & Shiang, T. Y. (2015). Assessment of walking, running, and jumping movement features by using the inertial measurement unit. *Gait Posture, 41*(4), 877-881. doi:10.1016/j.gaitpost.2015.03.007

Lin, K. C., Hung, H. C., & Young, S. S. C. (2014). The Effectiveness of adopting tablet PC in physical education of badminton skill. *Physical Education Journal, 47*(3), 437–449. doi:10.6222/pej.4703.201409.1009

Magogwe, J. M., Ntereke, B., & Phetlhe, K. R. (2014). Facebook and classroom group work: A Trial study involving university of Botswana advanced oral presentation students. *British Journal of Educational Technology, 46*(6), 1312-1323. doi:10.1111/bjet.12204

Manca, S., & Ranieri, M. (2013). Is it a tool suitable for learning? A Critical review of the literature on Facebook as a technology-enhanced learning environment. *Journal of Computer Assisted Learning, 29*(6), 487–504. doi:10.1111/jcal.12007

Manca, S., & Ranieri, M. (2016). Facebook and the others. Potentials and obstacles of social media for teaching in higher education. *Computers & Education, 95*, 216–230. doi:10.1016/j.compedu.2016.01.012

Mazman, S. G., & Usluel, Y. K. (2010). Modeling educational usage of Facebook. *Computers & Education, 55*(2), 444–453. doi:10.1016/j.compedu.2010.02.008

McKenzie, T. L., & Lounsbery, M. A. F. (2013). Physical education teacher effectiveness in a public health context. *Research Quarterly for Exercise and Sport, 84*(4), 419–430. doi:10.1080/02701367.2013.844025

Mcloughlin, C., & Lee, M. J. W. (2007). Social software and participatory learning: Pedagogical choices with technology affordances in the Web 2.0 era. In R. Atkinson, on, C. McBeath, S-K. A. Soong, & C. Cheers (Eds.), *ICT: Providing choices for learners and learning* (pp. 664–675). Singapore, Singapore: Centre for Educational Development, Nanyang Technological University.

Melograno, V. J. (2000). Designing a portfolio system for K-12 physical education: A Step-by-step process. Measurement in Physical Education and Exercise Science, 4(2), 97–115. doi:10.1207/S15327841Mpee0402_5

Miron, E., & Ravid, G. (2015). Facebook Groups as an academic teaching aid: Case study and recommendations for educators. *Educational Technology & Society, 18*(4), 371–384.

Newell, K. M. (1991). Motor skill acquisition. *Annual Review of Psychology, 42*, 213–237.

Papastergiou, M., & Gerodimos, V. (2013). Can learning of basketball be enhanced through a web-based multimedia course? An experimental study. *Education and Information Technologies, 18*(3), 459–478. doi:10.1007/s10639-012-9186-z

Pang, Y. (2010). Improving Hybrid Learning of Physical Education by Video Review. In X. Luo, M. Spaniol, L. Wang, Q. Li, W. Nejdl, & W. Zhang (Eds.), *Advances in Web-Based Learning – ICWL 2010 SE - 24* (pp. 230–239). doi:10.1007/978-3-642-17407-0_24

Potdevin, F., Vors, O., Huchez, A., Lamour, M., Davids, K., & Schnitzler, C. (2018). How can video feedback be used in physical education to support novice learning in gymnastics? Effects on motor learning, self-assessment and motivation. *Physical Education and Sport Pedagogy, 23*(6), 559–574. doi:10.1080/17408989.2018.1485138

Raza, S. A., Qazi, W., & Umer, A. (2017). Facebook is a source of social capital building among university students: Evidence from a developing country. *Journal of Educational Computing Research, 55*(3), 295–322. doi:10.1177/0735633116667357

Roberts, P., Maor, D., & Herrington, J. (2016). ePortfolio-Based learning environments: Recommendations for effective scaffolding of reflective thinking in higher education. *Educational Technology & Society, 19*(4), 22–33.

Roblyer, M. D., McDaniel, M., Webb, M., Herman, J., & Witty, J. V. (2010). Findings on Facebook in higher education: A Comparison of college faculty and student uses and perceptions of social networking sites. *The Internet and Higher Education, 13*(3), 134–140. doi:10.1016/j.iheduc.2010.03.002

Schmidt, R. A., & Lee, T. D. (1999). *Motor control and learning: A Behavioral emphasis.* Champaign, IL: Human Kinetics.

Statista. (2020). *Number of social media users worldwide from 2017 to 2025 (in billions).* Retrieved from https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

Statista. (2020). *Number of monthly active Facebook users worldwide as of 2nd quarter 2020 (in millions).* Retrieved from https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

Tsai, P. L., Liao, H. F., & Chen, C. Y. (2016). The Technical analysis of women's doubles in 2011 London world badminton championships. *NTU Journal of Physical Education, 31*, 1-8. doi:10.6569/NTUJPE.2016.31.01

Tolgfors, B. (2018). Different versions of assessment for learning in the subject of physical education. *Physical Education and Sport Pedagogy, 23*(3), 311–327. doi:10.1080/17408989.2018.1429589

Wise, L. Z., Skues, J., & Williams, B. (2011). Facebook in higher education promotes social but not academic engagement. In *Proceedings ASCILITE 2011* (pp. 1332–1342). Retrieved from https://www.ascilite.org/conferences/hobart11/downloads/ProceedingsV3.pdf

Yang, K., Li, J., & Xu, M. (2017). Artificial Intelligence computer assisted instruction system for basketball training (ICAI) characteristics and performance analysis. *Electronic Devices, 6*(1), 12-17.

Yu, C. H., Wu, C. C., Wang, J. S., Chen, H. Y., & Lin, Y. T. (2020). Learning tennis through video-based reflective learning by using motion-tracking sensors. *Educational Technology & Society, 23*(1), 64-77.

Zou, J., Liu, Q., & Yang, Z. (2012). Development of a Moodle course for schoolchildren's table tennis learning based on competence motivation theory: Its effectiveness in comparison to traditional training method. *Computers & Education, 59*(2), 294–303. doi:10.1016/j.compedu.2012.01.008

Goglio, V., & Parigi, P. (2020). The Social Dimension of Participation and Completion in MOOCs. *Educational Technology & Society*, *23* (4), 106–123.

The Social Dimension of Participation and Completion in MOOCs

Valentina Goglio[1,2*] and Paolo Parigi[2,3]
[1]University of Turin, Italy // [2]Stanford University, Stanford, CA, USA // [3]Facebook, Menlo Park, CA, USA //
valentina.goglio@unito.it // pparigi@stanford.edu
[*]Corresponding author

(Submitted April 14, 2020; Revised June 14, 2020; Accepted October 3, 2020)

ABSTRACT: The rapid and impressive development of Massive Open Online Courses (MOOCs) in less than half a decade has generated contrasting arguments about their social dimension. This paper investigates how the socio-economic background of learners affects their own experience and chances of course completion. The analyses test whether learners with a low socio-economic status (SES) have fewer chances of completing the online course and whether participation in online discussion forums moderates the role of SES. The data comes from two MOOCs provided by Stanford University. We find a negative association between low SES, course completion and course engagement. Moreover, we find that forum participation plays an ambiguous role, reinforcing the advantage of well-educated learners enrolled in one course, while it has no significant effect on the other course. The article concludes with some policy implications on social stratification in MOOCs and with some design suggestions for creators of MOOCs.

Keywords: MOOCs, SES, Social inequality, Higher education

1. Introduction

Between 2008 and 2012, the outburst of Massive Open Online Courses (MOOCs) in the USA created great excitement about their potential, putting them among the key challenges for the future of higher education. The hype that characterized the earlier period seems to be gone. MOOCs no longer enjoy the same attention in the media that they did in 2012-2013, however the number of users registered on the main MOOCs platforms is impressively high and keeps growing (Shah, 2019). This phenomenon has generated contrasting arguments (Sharrock, 2015). On one hand, these resources are enthusiastically welcomed as opening a series of opportunities for reducing social inequalities and promoting growth and employment. On the other, skeptics question the real empowering and equalizing effect of MOOCs.

Previous research has highlighted that MOOCs tend to attract already advantaged individuals. MOOC participants tend to be well-educated individuals, living in developed, mainly English-speaking countries (Emanuel, 2013). Highly educated learners are those with higher chances of completing the course, while learners from the Global South or countries with a medium to low Human Development Index (HDI), and women, tend to be disadvantaged in terms of persistence, completion and grades (Kizilcec & Halawa, 2015). Even within developed countries like the USA, analysis of the geolocation of learners based on zip codes showed that, although diffusion is quite widespread, the largest concentration of learners comes from wealthy neighborhoods (Glass et al., 2016). Yet, less advantaged individuals may have higher returns from MOOC attendance and completion (Hout, 2012), associated with lower costs and access barriers compared to formal education. Key factors which -in principle- may incentivize their persistence and completion.

As highlighted by van de Oudeweetering and Agirdag (2018a) in their systematic review of the literature, there is no univocal evidence about the social implications of MOOCs for underprivileged learners. On one hand, the empirical studies considered show an advantage for highly educated and employed individuals. However, not all proxies for the socio-economic status of learners are taken into consideration (e.g.: occupation is often overlooked) and results in terms of completion and engagement provide mixed evidence, suggesting that MOOCs may serve particular underprivileged groups in unforeseen ways.

The article aims to make a contribution in this direction, testing the association between some socio-economic characteristics and outcomes in terms of course engagement and participation. We consider two possible outcomes for MOOC learners: (i) completion, the attainment of a certificate of accomplishment; and (ii) engagement, defined as having solved a certain number of problems (setting the threshold at 60%). Moreover, the research design includes learner participation, operationalized as participation in an online forum by posting messages. The aim is to test whether the relationship between proxies for SES and outcome variables is moderated by participation in online forum discussions.

The empirical analysis proposed is relevant for two reasons. First, the high and growing number of learners enrolled in MOOCs means that these courses became a "common event" for many different people. Such widespread coverage makes it a socially relevant fact that deserves further attention with regard to its social implications, not only to pedagogical aspects associated with the learning experience. Second, the pervasive rhetoric of equalization and democratization of access to education that accompanied the diffusion of MOOCs is now being reconsidered and this work helps shed light on whether the association between SES and educational achievement –demonstrated by social stratification research- holds true also in the particular domain of MOOCs.

2. Theoretical framework

2.1. Inequality of educational opportunities

A long tradition of studies has proven the link between social origin, access and attainment of higher education, and the key role played by cultural capital and economic resources of the family of origin in shaping the different chances of children attaining post-secondary education. Children of families with high socio-economic status (measured as economic resources, educational titled or cultural capital) show higher attainment levels, higher grades and higher chances of attaining post-secondary education. The seminal work by Blau and Duncan (1967) was the first to empirically address this relationship, demonstrating the extent to which family background influences college attainment. Educational transition model literature (Mare, 1981) showed that the advantage of high socio-economic origin and its social-psychological benefits becomes particularly important for the highest transitions (i.e., college). Subsequent works further revealed how the socio-economic background of the family is one of the strongest predictors of college completion (though not the only one) (Roksa et al., 2007; Rosenbaum et al., 2009). Several studies investigated such association in a comparative perspective: the seminal work by Shavit and Blossfeld (1993) confirmed the stability of the effect of social origin on educational transitions, with their findings being subsequently contested by further analysis showing equalization trends in European countries.

The process of educational expansion that occurred in the second half of the 20th century seems to have contributed to reducing the relative inequality in access and attainment of higher education in the USA and in Europe (Bernardi & Ballarino, 2014; Breen, 2010; Breen & Jonsson, 2005). However, this process is only partial and a better examination of different measures of inequality and an encompassing analysis of dynamics throughout the whole 20th century reveals a more complex picture. Indeed, the process of expansion of access that characterized the higher education system in the USA after World War II, accompanied by public policy intervention, expanded enrollments. Yet, it was also accompanied by an increasing stratification of the system, with expansion occurring mainly at low-status institutions while the advantage associated to higher social status did not decrease (Roksa et al., 2007). A recent analysis of the dynamics of educational inequality in the USA envisages a possible increase in inequality of educational attainment for younger people, born after 1990, who experience the combination of decreasing financial support for education and increasing wage inequalities (Bernardi, Hertel, & Yastrebov, 2018).

Comparative analysis shows that this trend is accompanied by processes of horizontal stratification (Shavit et al., 2007): according to the "effectively maintained inequality" approach (Lucas 2001), parental education still plays a critical role in terms of quality, reputation and duration of the course attended, even in terms of prestige of the field of study. In a context of expansion of access to education, more affluent families mobilize their resources to secure educational credentials that are qualitatively superior and more prestigious in order to guarantee better outcomes to their offspring (Gerber & Cheung, 2008). Empirical research on European countries confirms that individuals with culturally advantaged parents are more likely to attain more rewarding, longer and prestigious educational qualifications from high quality institutions (Triventi, 2013; Triventi et al., 2017). We have summarized this literature with the following hypothesis:

H1: We expect to find a negative association between low SES and completion of the MOOC, represented by the attainment of a certificate of accomplishment (outcome 1)

2.2. Completion and participation in MOOCs

The definition of success in particular types of online courses like MOOCs is still subject to discussion by academics and the general public (Ho et al., 2015). The main objection to the attainment of a certificate of completion in MOOCs as a measure of success is that it may not have the same informative value that it has in the domain of traditional education (Evans & Baker, 2016a; Evans et al., 2016b; Kizilcec et al., 2013). Yet, consensus over alternative definitions and measurement of success in MOOCs has not been reached yet. Evans and Baker (2016a) compare different indicators of persistence besides the attainment of a certificate of completion and propose more inclusive measures of "success" in MOOCs, such as the cumulative grade reached by the learner or the percentage of videos watched, and an associated threshold for achievement. They identify three different outcome variables of interest: (i) engagement; (ii) persistence; (iii) completion, and analyze the factors at student, lecture and course-level affecting the results. Kizilcec, Piech, and Schneider (2013) draw a typology of MOOC learners and identify four types, pointing to the different uses that learners make of MOOCs, and to a more complex understanding of the multifaceted group of those who do not complete all activities.

Despite contrasting views, the certificate of completion and the number of activities completed during the course is still considered a valuable piece of information about learners' behavior and success (Almeda et al., 2018; Reich, 2014; Wang & Baker, 2018).

H2: we expect to find a negative association between low SES and engagement in the MOOC, defined as having taken at least 60% of the quizzes available (outcome 2)

Most MOOCs also include an online discussion forum where learners can interact through forum posts, starting discussions on the class's topic autonomously, replying to discussions or assignments launched by the teacher or seeking advice on course material among peers or from the teacher. This feature represents an attempt to address student and teacher frustration due to the lack of social interaction typical of the physical classroom (Beard et al., 2004). Participation in online forum discussions can represent a substitute for classroom interaction and, despite their limitations and distinctive features, it can be argued that they tend to reproduce some of the social stratification mechanisms typical of face-to-face interaction in formal schooling.

Empirical literature on the type of participation in MOOC online forums and the social characteristics of people engaged in this activity is still limited. Most of the research available points to a positive association between participation in forum discussions and positive outcomes in MOOCs (Almeda et al., 2018; Zhang et al., 2017). However, the empirical findings on the social and demographic characteristics of learners posting on the MOOC forum are limited and often contradictory (Ruthotto et al., 2020). Research into traditional classroom interaction showed that participation is positively correlated to learning outcomes (Rocca, 2010) but is not evenly distributed across social groups and can be challenging for certain categories of learners, such as students from lower SES or minority groups (Chang, 2005). In this respect, the Bourdieu concept of cultural capital can prove helpful: interacting with other students or the teacher requires cognitive skills, competence, social abilities and specific attitudes toward education. These represent primary effects of the class of origin, an incorporated form of cultural capital transmitted from high social status families to their offspring (Sallaz & Zavisca, 2007; Schizzerotto & Barone, 2012). This ultimately turns into a cycle that reinforces inequality for highly educated individuals who already possess good skills for succeeding in learning, but are also further advantaged by their ability to interact and benefit from this interaction with others in the online discussion forum, resulting in better learning outcomes.

Empirical research into interaction through digital media also highlights the fact that anonymity, missing social clues and lower emotional involvement in computer-mediated communication may contribute to the creation of a neutral and more inclusive environment, fostering participation by all (Gunn et al., 2003; Kollock & Smith, 1996). This suggests a *technicist* hypothesis in which the particular context of the online discussion forum of MOOCs may activate traditionally less active students. The anonymity granted by missing social clues in the online discussion forum may act as a digital filter, reducing the stress associated with exposing personal thoughts to the judgment of others. In the case of MOOCs, participation in online discussion forums may positively moderate the relationship between socio-economic characteristics of the learner and course completion and engagement:

H3: we expect to observe that participation in online forums positively moderates the negative effect of low SES on both completion and engagement. We expect to find that low SES learners who write at least one forum post have better chances of completing or staying engaged than their peers who do not participate in the forum

3. Material and methods

The empirical analyses presented in the paper use secondary data from two different MOOCs provided by Stanford University through their OpenEdX instance, Lagunita. The data include demographic, performance and survey-related information on learners registered in two MOOCs: "Statistical Learning" (hereafter SL) and "America's Poverty and Inequality" (hereafter API), released in 2016, respectively in the Winter and Fall term. The SL course is an introductory-level course in statistical modeling and data science, with a focus on regression and classification methods, developed by professors T. Hastie and R. Tibshirani. It consists of 9 modules (self-paced) and the workload is estimated at 3–5 hours per week.

The API course is developed by the Stanford Center on Poverty and Inequality and deals with income inequality, poverty, racial and gender inequalities in the USA. It is an introductory level course consisting of 9 modules (self-paced) and requires about 2 to 4 hours of study per week.

The two courses are not a representative sample of online classes but were chosen following the strategy of maximizing differences across courses available in Lagunita and assuring a sufficient sample size.

The demographic information and survey data come from an embedded pre-course survey, designed by the CAROL research team at Stanford University, administered to learners at the very beginning of the MOOC (after enrollment, before module 1). We are aware of issues associated with response errors in the MOOC surveys, leading to inaccurate representations of the population of learners and bias in estimating the effect of demographic variables. However, as indicated by van de Oudeweetering and Agirdag (2018b), embedded web surveys help increase the response rate (compared to email surveys) and help reach learners who are generally less-responsive, ultimately reducing response bias. Table 1 below supports this argument: among people who completed the course, the survey response rate of the is just below 50%, providing a solid source of data to describe dynamics among completers. Dropouts, i.e., those that did not complete the course and did not answer the survey, are outside the scope of this paper.

Table 1. Learners who completed the courses and survey respondents

Responded to pre-course survey	Completed course: Yes	Completed course: No	Total
Yes	48%	13%	10,787
No	52%	87%	62,067
Total	100% (N = 4,491)	100% (N = 68,363)	72,854

The empirical strategy uses quantitative methods of analysis based on logistic regression analysis to investigate the direct and indirect relationship between the socio-economic background of learners and MOOC completion and engagement. Models are run separately for the two courses, due to the different sample size and definition of the dependent and independent variables.

3.1. Dependent variables

In order to take into consideration different approaches to "success" in MOOCs, we considered two outcome variables. The first is "*completion*," defined as a dummy variable equal to 1 if the learner downloaded the *certificate of accomplishment* available at the end of the course. This may not be the perfect variable for completion, as not all learners may be interested in downloading the certificate at the end of the course. Yet, such information is available for both courses and is freely downloadable. Rules of attainment change slightly between the courses: learners must have earned at least 50% of graded assignments for SL, with at least 75% for the API course. This difference can lead to a more restricted sample of completers for the second course.

The second outcome variable is the *percentage of quizzes answered*, a proxy for *engagement*. The courses analyzed are organized in several chapters. Each includes a video lecture and an assessment in the form of a multiple-choice quiz (the proportion of video lectures watched was also considered but the two activities are highly correlated, about 0.75). The minimum level of engagement is coded as at least 60% of quizzes answered, irrespective of the score attained, out of the total number of quizzes available.

3.2. Independent variables

Typically, literature on social stratification refers to the level of education and occupational category of parents, as well as family and personal income or job prestige as proxies for SES (Breen & Jonnson, 2005; Ganzeboom et al., 1992; Shavit & Blossfeld, 1993). The data available for this study offer limited information on these indicators. Consequently, the study relies on a series of other variables that are only implicit indicators of the socio-economic background: (i) educational attainment of the learner; (ii) self-declared employment status; (iii) educational attainment of parents (for the API course only). Employment status can be considered an imperfect but pragmatic indicator of current socio-economic status of the individual when information on the type or prestige of the job is not available. Indeed, jobless individuals are a vulnerable group exposed to greater risk of poverty and social exclusion due to lack of economic resources from paid employment, but also to reduced or poor quality professional and social networks compared to employed learners (Gallie & Paugam, 1999). In addition, with regard to educational attainment, when the sample is made up of adult individuals the influence of parental background is still present but more distant (Kizilcec & Saltarelli 2019), making current individual status more relevant.

The moderating factor in the model - participation in the online discussion forum of the course - is defined as equal to 1 if the learner wrote at least one forum post after the second video watched; equal to 0 if the learner (i) did not write any post or (ii) wrote a post before watching the second video. In order to mitigate the risk of endogeneity, where participation is correlated with motivation, influencing the outcome variable, forum posts written before watching the second video of the course are excluded (alternative arrangements were also considered, but the relationship among variables did not change. This decision is based on the assumption that writing before the start of the course or at the very beginning (e.g., before the second video) may indicate extremely high levels of motivation of the learner (ultimately influencing the final outcome) and not actual requests for clarification or support on the study material. Finally, multicollinearity issues are excluded as VIF = 1.60 for SL and VIF = 2 for API.

4. Results and discussion

Table 2 provides a summary of the characteristics of learners in the two courses. The sample size of demographic and survey data varies widely between the two courses, with SL attracting a much higher number of learners than API (25 times more). The latter shows better engagement proxies with a higher proportion of: (i) respondents to the pre-course survey, (ii) certificates attained (despite the more restrictive rule for the Certificate of Accomplishment) and (iii) written online forum posts. Women are underrepresented in SL, while the distribution of educational qualifications is quite similar in both courses and oriented toward high levels of education, with a high proportion of learners with Master's Degrees. Employed and unemployed people are equally represented in both courses, while SL has a higher proportion of students; API has a higher proportion of retired people and a residual category of people not falling within any of the other standard employment categories (other).

Table 2. Sample characteristics

	Statistical learning		Poverty and inequality	
	N	*%*	*N*	*%*
Total N	72,854	100.00%	2,908	100.00%
	Dependent variables			
Took certificate of completion	4,491	6.16%	358	12.31%
Answered at least 60% of quizzes	4,256	5.84%	469	16.13%
	Independent variables			
Wrote at least one forum post	989	1.36%	397	13.65%
Wrote post & took certificate	499	0.68%	209	7.18%
	Demographic data			
Female	13,937	22.22%	1,609	55.33%
Educational attainment				
Less than BA	6,035	8.28%	298	10.25%
Bachelor's	23,481	32.23%	763	26.24%
Doctorate	7,155	9.82%	373	12.83%
Master's/prof. Degree	25,781	35.39%	1,173	40.34%
Withheld	9,180	12.60%	292	10.04%
Geographical distribution				
USA	24,988	34.30%	2,072	71.25%
Europe	12,955	17.78%	206	7.08%
India	9,001	12.35%	39	1.34%
China	2,412	3.31%	21	0.72%
	Survey data			

Total N	10,787	100.00%	938	100.00%
Employment status				
FT employed	6,619	64.52%	633	69.48%
PT employed	662	6.45%	n.a.	n.a.
Student	2,106	20.53%	85	9.33%
Unemployed	766	7.47%	50	5.49%
Retired	106	1.03%	88	9.66%
Other	n.a.	n.a.	55	6.04%
Parental education				
High school or less	n.a.	n.a.	270	29.54%
Associate/Some College	n.a.	n.a.	273	29.87%
Bachelor's	n.a.	n.a.	54	5.91%
Master's/PhD	n.a.	n.a.	317	34.68%
	Other independent variables			
	Mean	SD	Mean	SD
No. Posts written	3.18	6.39	6.21	13.23
Age	31.41	9.53	40.65	14.67

Note. Source: Own elaboration on Stanford Lagunita data.

4.1. Certificate of completion

Starting from the course in SL, Model 1 in Appendix A shows the odds ratio of the logistic regression model with outcome variable the attainment of a certificate of completion and proxies for SES as predictors. An odds ratio greater than 1 indicates a positive association, while an odds ratio lower than 1 indicates a negative association. Figure 1 shows the average marginal effects of SES proxies, indicating actual differences in the probability of attaining a certificate of completion for each group considered (Leeper, 2018). The left panel indicates the estimated average marginal effect for the level of education of learners (reference category: less than bachelor's degree); the right panel represents estimates for the employment situation of the learner (reference category: full time employment). Results for the former support the hypothesis of a relative disadvantage of learners with low educational qualifications: indeed, learners with a university education, in particular PhD (+9 p.p.) and, to a lesser extent, master's graduates (+3.6 p.p.) have a higher chance of completing the course, compared to their peers without a college degree.

As far as the employment situation is considered, the middle panel shows that unemployed individuals have a 5 p.p. higher chance of attaining a certificate of accomplishment compared to individuals who are in full time employment. This may suggest that unemployed learners may be more motivated towards gaining and updating skills and may view or use the certificate of completion as a signaling tool for increasing their employment prospects. Their potentially disadvantaged condition (see van de Oudeweetering & Agirdag, 2018a) seems to works as an incentive for an "instrumental" use of the course, referring to Max Weber's theory of social action (Weber, [1922] 1978). By completing the course and attaining a certificate, unemployed learners pursue a "goal rational" type of social action, aimed at acquiring new skills or updating existing ones and, at the same time, showing prospective employers their skills and willingness to continue learning (Castaño-Muñoz et al., 2017).

The model then tests whether participation in the online forum has a moderating role on the relationship between SES proxies and completion, indicating whether or not writing at least one forum post is associated with an advantage for certain groups (in the chances of completion with a certificate).

Table 3 shows the odds ratio of the regression model including the interaction effect between writing at least one forum post (after watching the second video) and each predictor. Writing at least one forum post has a positive and strong direct association with the chances of attaining a certificate of completion. However, when introducing the interaction between level of education and having written at least one post (Model 1), the estimates show that there is no significant interaction between level of education and the writing of at least one forum post, with the only exception being those with a PhD (90% confidence interval), who have an advantage. This advantage remains when considering the main effect of the predictors, indicating that, among those who wrote at least one forum post, those with a PhD are the group that benefits the most. Model 2, on the other hand, shows a negative interaction of writing a forum post with the condition of being a student. However, when considering the main effect of the two interacting predictors, students who wrote a post still have an advantage, compared to their peers who did not write any forum post.

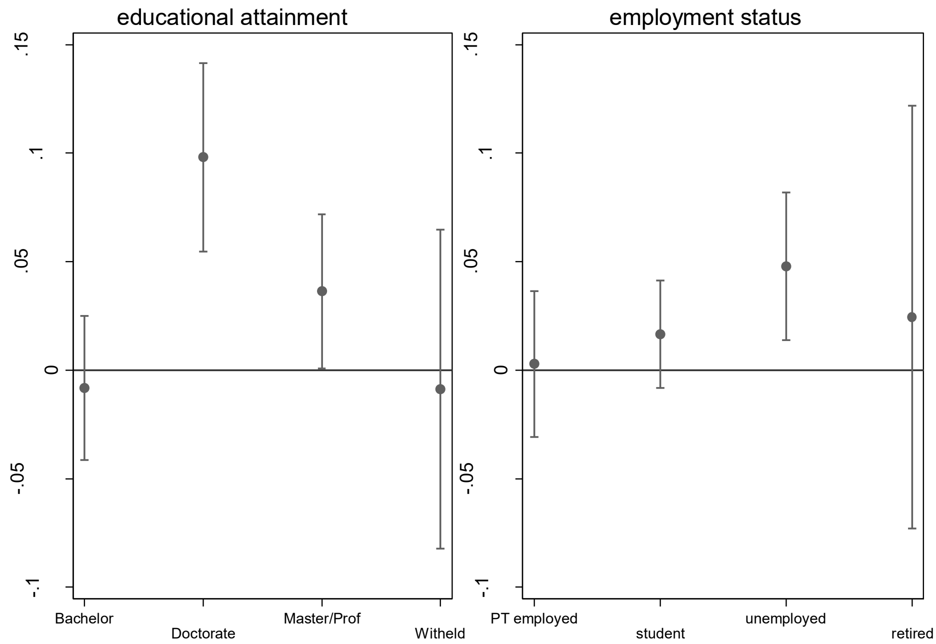


Figure 1. Statistical learning, average marginal effects for outcome = completion

Table 3. Statistical learning. Estimates for logistic regression models with interactions (odds ratios)

	Model 1	Model 2
	Outcome = certificate	Outcome = certificate
Employment status (ref = FT employed)		
PT employed	0.954	0.997
	(0.107)	(0.115)
Student	1.108	1.147
	(0.0864)	(0.0904)
Unemployed	1.178	1.214
	(0.119)	(0.127)
Retired	0.938	1.043
	(0.300)	(0.345)
Education (ref = less than BA)		
Bachelor’s	0.992	1.020
	(0.121)	(0.122)
Doctorate	1.806**	1.894**
	(0.257)	(0.264)
Master’s/professional degree	1.309*	1.344*
	(0.166)	(0.167)
Withheld	1.112	1.097
	(0.294)	(0.290)
Wrote at least one forum post (after the second video)	5.353**	13.28**
	(2.285)	(2.451)
Interaction education &post		
Bachelor’s & post	1.796	
	(0.886)	
Doctorate & post	2.738	
	(1.524)	
Master’s & post	1.733	
	(0.810)	
Withheld & post	Omitted	
Interaction employment &post		
PT employed & post		0.495
		(0.220)
Student & post		0.375**
		(0.137)
Unemployed & post		0.584

		(0.221)
Retired & post		0.437
		(0.335)
Sex	Y	Y
Age	Y	Y
Constant	0.183**	0.176**
	(0.0248)	(0.0236)
Observations	9,111	9,112

Note. Standard errors in parentheses; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

With respect to the second course analyzed, API, Figure 2 (and model 1 in Appendix B) shows the association between the selected proxies for socio-economic background of the learner and the chances of attaining a certificate of completion (awarded after 75% of correct answers in this course). Overall, the estimates seem to support our hypothesis of a positive association between higher SES and the chance of completion with a certificate, with some peculiar features. Indeed, contrary to that found for the other course, in this case there are no significant differences based on the educational level of the learner (left panel of Figure 2). When it comes to the employment status of the learner, unlike that observed for the other course, people involuntarily excluded from the labor market do not have a different chance of completing the course, compared to their employed peers. It is a residual category, individuals self-defined as having “other” employment status (not employed, nor unemployed or students), that shows a higher chance (+22 p.p.) of completing the course with a certificate. Finally, the educational background of parents (available for this course only) does not show a robust association with the outcome variable, with the exception of a slightly significant advantage (90% confidence interval) of learners with at least one parent with a university education (+7 p.p. higher chance of completion with a certificate for offspring of parents with a Master’s degree or PhD).

The findings related to the residual category of “other” labor market status are unexpected and may indicate a particular group of completely inactive individuals. On one hand, they may fall within the definition of NEETs (not in education, employment or training). However, upon closer examination, individuals in this group tend to be adult learners (mean age = 43.9, median age = 46) and two-thirds are female learners (67%). This background could indicate that those learners may potentially come from better socio-economic backgrounds, as they can afford to stay out of the labor market (are not employed) without seeking employment (they are not unemployed), nor being in formal education (they are not students). In addition, taking into consideration the fact that (i) the definition of NEETs in Europe includes young people up to the age of 34 (Eurostat, 2019), while the sample here has a much higher average age; (ii) that 71% of the learners of this course come from the USA; (iii) that, in the USA, figures available for NEETs are significantly lower than in some European countries (OECD, 2018), we can argue that this category could represent an advantaged segment of MOOC learners.

Continuing with the reference to Max Weber’s theory of action, in this case it can be argued that learners on the API course may follow a “value rational” type of social action, based on shared values on an important social issue. The instrumental value attributed to the SL course seems to fail in favor of a more value-oriented approach or recreational function of the course.

Table 4 shows estimates of the interaction between the predictors and participation in an online forum discussion. Model 1 confirms that employment status, and in particular being in the category of “other” is the strongest predictor and the fact that having written a post does not show any positive association either when considered alone or interacting with educational level. Model 2 shows no significant interaction between forum participation and different labor market situations, indicating that writing a forum post does not lead to any different outcome depending on the labor market situation. Model 3 shows that writing a forum post is positively associated with a higher chance of completion with a certificate and a slightly significant advantage of learners coming from highly educated parents who write a post. Among those who write at least one forum post after the second video, learners from highly educated parents have a higher chance of obtaining a certificate of completion compared to learners who participate but are from low educated parents, with a 90% confidence interval.

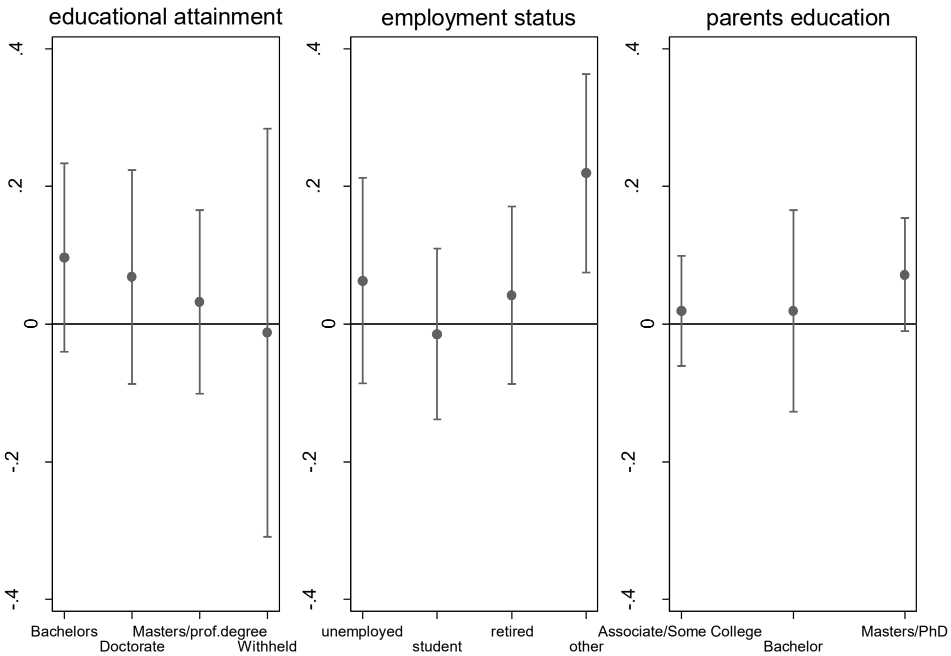


Figure 2. America's poverty and inequality, average marginal effects for outcome = completion

Table 4. America's poverty & inequality. Estimates for logistic regression model with interaction (odds ratio)

	(1) Outcome = certificate	(2) Outcome = certificate	(3) Outcome = certificate
Employment status (ref = FT employed)			
Unemployed	1.282	1.454	1.284
	(0.444)	(0.569)	(0.448)
Student	0.890	0.877	0.850
	(0.281)	(0.317)	(0.269)
Retired	1.171	1.434	1.154
	(0.359)	(0.482)	(0.351)
Other	2.541**	2.359*	2.381**
	(0.818)	(0.872)	(0.757)
Education (ref = associate degree or less)			
Bachelor's	1.276	1.638	1.646
	(0.500)	(0.582)	(0.579)
Doctorate	1.140	1.484	1.420
	(0.502)	(0.595)	(0.565)
Master's/prof. degree	0.928	1.276	1.275
	(0.359)	(0.454)	(0.447)
Withheld	1.307	1.155	1.175
	(1.067)	(0.945)	(0.949)
Parental education (ref= high school or less)			
Associate/Some College	1.143	1.117	1.006
	(0.232)	(0.226)	(0.231)
Bachelor's	1.110	1.110	0.913
	(0.405)	(0.405)	(0.379)
Master's/PhD	1.395	1.407	1.142
	(0.284)	(0.285)	(0.262)
Wrote at least one forum post (after the 2nd video)	1.170	3.546**	2.017*
Interaction education & post	(0.808)	(0.795)	(0.644)
Bachelor's & post	2.499		
	(1.899)		
Doctorate & post	2.647		
	(2.220)		
Master's/prof.degree & post	3.382		

	(2.515)		
Interaction employment & post			
Unemployed & post		0.613	
		(0.497)	
Student & post		0.823	
		(0.541)	
Retired & post		0.485	
		(0.265)	
Other & post		0.979	
		(0.713)	
Interaction parental edu & post			
Associate/Some College & post			1.497
			(0.691)
Bachelor's & post			2.392
			(2.140)
Master's/PhD & post			2.217
			(1.004)
Sex	Y	Y	Y
Age	Y	Y	Y
Constant	0.396*	0.305**	0.347*
	(0.173)	(0.128)	(0.146)
Observations	823	824	824

Note. Standard errors in parentheses; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

4.2. Engagement

With respect to the first course analyzed, SL, Figure 3 shows the average marginal effect associated with two proxies for socio-economic background on the chances of responding to at least 60% of quizzes (regardless of the grade received) (for coefficients see model 2 in Appendix A). The left panel shows that the level of education is the only predictor showing a close and significant association: learners with a PhD or master's degree have a higher chance (+11 p.p. and +5 p.p.) of staying engaged throughout the course compared to their peers with less than a bachelor's degree, and values are similar to the case of course completion. The right panel on the other hand, shows that unemployed learners and students have a similar (limited) higher likelihood of staying engaged (+4 p.p. and +3 p.p.), although statistically significant at a 90% confidence interval only. Consequently, the advantage of unemployed learners observed for the attainment of a certificate of completion is no longer robust when considering the likelihood of taking at least 60% of the quizzes.

As far as the interaction with participation in forum discussions is considered, Table 5 shows the odds ratio of the interaction between participation in forum discussions and SES proxies. Neither model shows any significant interaction, indicating that writing a post is associated with a higher chance of engagement (main effect), but there are no significant differences across groups of learners (interactions), suggesting that participation in forum discussions does not play a moderating role between socio-economic status and engagement throughout the course.

Results for the API course are shown in Figure 4 (see model 2 in Appendix B for coefficients). The estimates for the likelihood of remaining engaged in the course reveal a slightly different situation compared to the attainment of a certificate.

As already observed, different educational levels (of the learner) are not associated with a higher or lower likelihood of staying engaged beyond 60% of course quizzes (left panel Figure 4). When looking at the employment status, the relative advantage of the residual category of the "other" labor market situation is confirmed (+16 p.p.), but this time together with a relative advantage of unemployed learners. Unlike the case of course completion, in the case of engagement, unemployed learners show a significantly higher likelihood of staying engaged throughout the course compared to employed individuals (+20 p.p.). Finally, social origin - proxied by parental education- is not robust and significantly associated with different likelihoods of engagement in the course material, with the exception of a small and slightly significant advantage for children of highly educated parents.

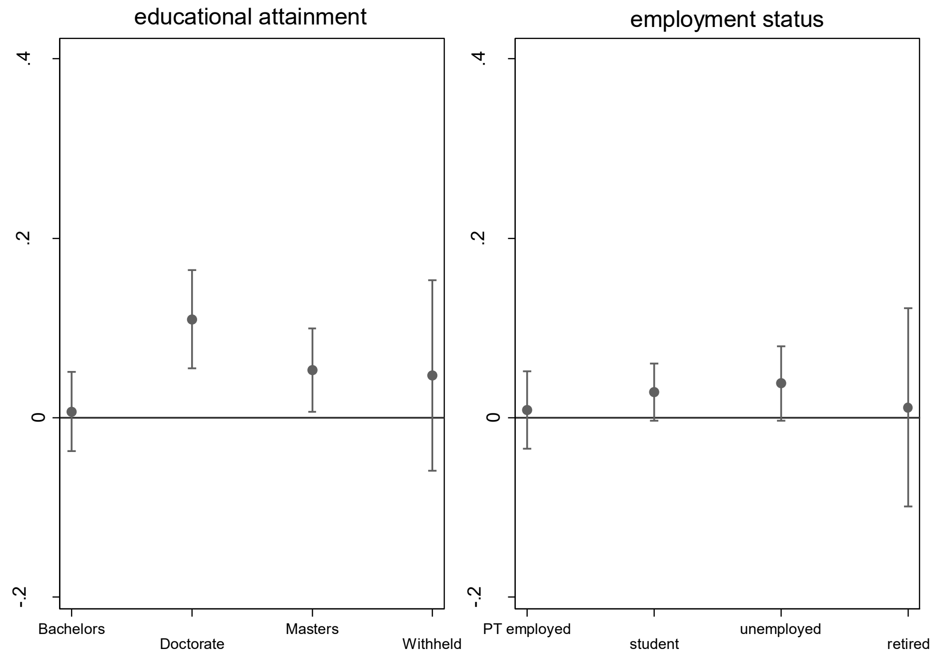


Figure 3. Statistical learning, average marginal effects for outcome = engagement

Table 5. Statistical learning. Estimates for logistic regression models with interactions (odds ratios)

	Model 1	Model 2
	Outcome=engagement	Outcome=engagement
Employment status (ref = FT employed)		
PT employed	1.015	1.052
	(0.113)	(0.120)
Student	1.147	1.147
	(0.0898)	(0.0908)
Unemployed	1.068	1.104
	(0.110)	(0.117)
Retired	0.921	0.961
	(0.270)	(0.294)
Education (ref = associate degree or less)		
Bachelor's	1.056	1.074
	(0.133)	(0.132)
Doctorate	1.854**	1.862**
	(0.269)	(0.265)
Master's/prof. Degree	1.377*	1.386*
	(0.179)	(0.177)
Withheld	1.540	1.566
	(0.411)	(0.414)
Wrote at least one forum post (after the second video)	6.473**	8.307**
	(2.897)	(1.544)
Interaction education & forum post		
Bachelor's & post	1.384	
	(0.720)	
Doctorate & post	0.961	
	(0.531)	
Master's/prof.degree & post	1.038	
	(0.508)	
Withheld & post	Omitted	
Interaction employment & post		
PT employed & post		0.532
		(0.240)
Student & post		1.058

		(0.436)
Unemployed & post		0.633
		(0.241)
Retired & post		0.604
		(0.471)
Sex	Y	Y
Age	Y	Y
Constant	0.290**	0.286**
	(0.0405)	(0.0393)
Observations	7,171	7,172

Note. Standard errors in parentheses; *$p < .05$; **$p < .01$; ***$p < .001$.

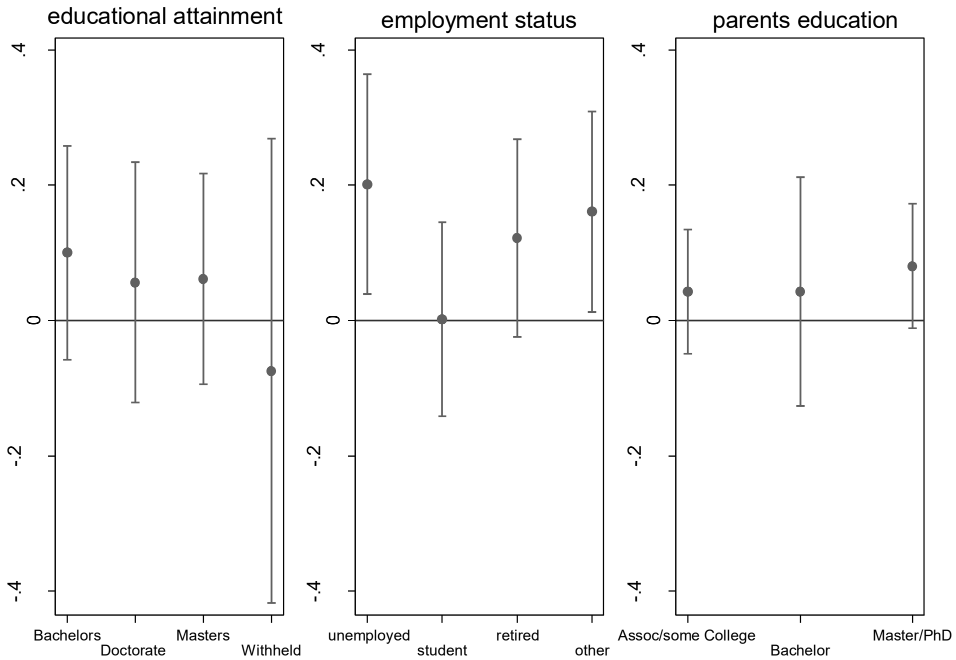


Figure 4. America's poverty & inequality, average marginal effects for outcome = engagement

Table 6. America's poverty & inequality. Estimates from logistic regression including interactions (odds ratio)

	(1) Outcome= engagement	(2) Outcome= engagement	(3) Outcome= engagement
Employment status (ref = FT employed)			
Unemployed	2.255*	2.017	2.262*
	(0.860)	(0.850)	(0.866)
Student	0.910	0.831	0.916
	(0.291)	(0.302)	(0.291)
Retired	1.664	1.847	1.617
	(0.533)	(0.650)	(0.515)
Other	1.975*	1.474	1.804
	(0.668)	(0.562)	(0.601)
Education (ref = associate degree or less)			
Bachelor's	1.253	1.591	1.560
	(0.501)	(0.581)	(0.561)
Doctorate	1.127	1.328	1.283
	(0.505)	(0.543)	(0.520)
Master's/prof. Degree	1.161	1.409	1.359
	(0.457)	(0.511)	(0.484)
Withheld	1.106	0.796	0.827
	(0.970)	(0.697)	(0.701)
Parental education (ref = high school or less)			
Associate/some college	1.207	1.217	1.167

	(0.249)	(0.251)	(0.269)
Bachelor’s	1.138	1.202	1.167
	(0.434)	(0.457)	(0.495)
Master’s/phd	1.382	1.395	1.247
	(0.288)	(0.292)	(0.289)
Wrote at least one forum post (after the 2nd video)	1.385	2.772**	2.380*
	(1.069)	(0.656)	(0.824)
Interaction education & post			
Bachelor’s & post	2.655		
	(2.258)		
Doctorate & post	2.106		
	(1.923)		
Master’s/prof.degree & post	2.160		
	(1.779)		
Interaction employment & post			
Unemployed & post		2.247	
		(2.620)	
Student & post		1.353	
		(0.977)	
Retired & post		0.644	
		(0.381)	
Other & post		2.626	
		(2.339)	
Interaction parental edu & post			
Associate/Some College & post			1.142
			(0.562)
Bachelor’s & post			1.045
			(0.950)
Master’s/phd & post			1.681
			(0.833)
Sex	Y	Y	Y
Age	Y	Y	Y
Constant	0.554	0.477	0.500
	(0.244)	(0.201)	(0.211)
Observations	710	711	711

Note. Standard errors in parentheses; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

With respect to the moderating effect of participation in forum discussions (Table 6), the estimates do not report any significant interaction of participation in forum discussions and the main predictors of SES. As for the case of attainment of a certificate of completion, participating in the forum discussion does not moderate the main effect of the SES predictors considered.

Consequently, findings for the API course suggest that availability of time is a crucial resource in determining the chances of staying engaged (unemployed and inactive learners are those who presumably have greater availability of time), while participation to forum discussions does not impact differently on different groups of learners.

5. Conclusions

This study analyzed the relationship between the socio-economic background of learners, their engagement, completion and participation in MOOCs. The main aim was to test whether some of the forms of social inequalities observed in the domain of traditional formal education apply also to the digital environment of MOOCs, but also whether MOOCs may, to some extent, help reach certain underprivileged groups in unforeseen ways.

Overall, findings support the main hypotheses of this research. Estimates from logistic regressions on the chances of attaining a certificate of completion and remaining engaged for at least 60% of the course material support the hypothesis of a negative association between low SES and outcome variables, though with some peculiarities.

In line with the hypothesis formulated, results for the SL course show that learners with a university education have a higher chance of completing as well as staying engaged in course materials. Contrary to that hypothesized, unemployed learners also show a higher chance of attaining a certificate of completion (compared to employed peers) and, to a lesser extent, of remaining engaged in the course activities.

Findings for the API course show the relative advantage of a particular group of learners: completely inactive people who are not employed, not seeking job and not in formal education. This group, mainly consisting of adult female learners, shows a higher chance of completing and staying engaged in the course. We speculate that this may represent a relatively advantaged group, made up of individuals who can afford to remain outside the labor market, without working and without seeking employment, supporting the idea that wealthier individuals tend to have a higher chance of achieving positive outcomes in MOOCs. When considering the alternative outcome variable, "engagement," unemployed people also show a higher chance compared to employed peers. This seems to suggest that greater availability of time can be a crucial factor for staying engaged (both inactive and unemployed learners are likely to have more time than their peers) but, for unemployed people, not enough to attain a certificate.

Moreover, findings seem to indicate that learners may attribute different uses to the courses considered. Following the terminology of Max Weber's theory of social action, we distinguished between "goal-rational" and "value-rational" types of social action in MOOCs. The higher chances of completion shown by people excluded from the labor market in SL could indicate that learners may rationally decide to complete the course in order to improve their employment opportunities, based on the expectation that prospective employers may appreciate this as a signal of their competences and willingness to reskill (or upskill). On the contrary, API learners may act on the basis of value-rational considerations, based on shared concerns relating to social problems.

Finally, our analysis shows an ambiguous moderating role of the participation in online discussion forums. Participation in an online forum does not change the original advantage of certain groups; rather, when slightly significant, it tends to reinforce this advantage. This finding does not support the "technicist" hypothesis attributing a positive moderating effect of the online discussion forum. On the contrary, findings provide support for the alternative explanation associated with the Bourdieu concept of cultural capital. In this framework, higher cultural and cognitive skills of people with higher educational levels represent a further advantage: make them more confident and more prone to interact in the online discussion forum, ultimately further advantaging them in the learning process, delineating a typical "Mathew effect" (see Merton, 1968).

Our findings have implications on course design. If participation does not moderate or if, when it does, it tends to reinforce the advantage of well-educated people or those who can afford to remain outside the labor market, interventions in the course design targeted at overcoming barriers for all learners may contribute to reshaping the trend. Participation in forums may be incentivized and targeted toward certain goals by: (i) making it a substantial part of the total grading; (ii) making it more attractive and user-friendly; (iii) providing regular message alerts; (iv) regulating the maximum length or number of posts per learner (in order to avoid "superposters") or even (v) making it mandatory for attaining the certificate.

This study also presents some limitations that future empirical research needs to explore: the limited number of courses analyzed and the lack of proper measures for assessing the socio-economic background of learners. Nevertheless, this study significantly contributes to addressing a pragmatic question: "While MOOCs may not be ideal, can they at least represent a viable option for some learners and in some circumstances?" (Literat, 2015, p. 1173). The higher likelihood that unemployed individuals will complete the course with a certificate, and the arguable differential value of the two courses suggest that MOOCs may serve different needs, emerging from different types of user, attributing various functions to a flexible educational tool like the MOOC.

Acknowledgement

The authors thank the Center for Advanced Research through Online Learning (CAROL) at Stanford University for the provision of data and Mitchell Stevens for his support and collaboration during the project.

This project has received funding from the H2020-MSCA-IF-2016 program under grant agreement no. 750242. The work reflects only the author's view and the Commission is not responsible for any use that may be made of the information it contains.

References

Almeda, M. V., Zuech, J., Baker, R. S., Utz, C., Higgins, G., & Reynolds, R. (2018). Comparing the factors that predict completion and grades among for-credit and open/MOOC students in online learning. *Online Learning, 22*(1), 1-18. doi:10.24059/olj.v22i1.1060

Beard, L. A., Harper, C., & Riley, G. (2004). Online versus on-campus instruction: Student attitudes & perceptions. *TechTrends, 48*(6), 29–31. doi:10.1007/BF02763579

Bernardi, F., & Ballarino, G. (2014). Participation, equality of opportunity and returns to tertiary education in contemporary Europe. *European Societies, 16*(3), 422–442. doi:10.1080/14616696.2012.750729

Bernardi, F., Hertel, F. R., & Yastrebov, G. (2018). A U-turn in inequality in college attainment by parental education in the US? *Research in Social Stratification and Mobility 58,* 33-43. doi:10.1016/j.rssm.2018.09.003

Blau, P. M., & Duncan, O. D. (1967). *The American occupational structure.* New York, NY: John Wiley & Sons.

Breen, R. (2010). Educational expansion and social mobility in the 20th Century. *Social Forces, 89*(2), 365–388. doi:10.1353/sof.2010.0076

Breen, R., & Jonsson, J. O. (2005). Inequality of opportunity in comparative perspective: Recent research on educational attainment and social mobility. *Annual Review of Sociology, 31*(1), 223–243. doi:10.1146/annurev.soc.31.041304.122232

Castaño-Muñoz, J., Kreijns, K., Kalz, M., & Punie, Y. (2017). Does digital competence and occupational setting influence MOOC participation? Evidence from a cross-course survey. *Journal of Computing in Higher Education, 29*(1), 28–46. doi:10.1007/s12528-016-9123-z

Chang, J. C. (2005). Faculty-student interaction at the community college: A Focus on students of color. *Research in Higher Education, 46*(7), 769–802.

Emanuel, E. J. (2013). Online education: MOOCs taken by educated few. *Nature, 503*(7476), 342–342. doi:10.1038/503342a

Evans, B. J., & Baker, R. B. (2016a). MOOCs and persistence: Definitions and predictors. In S. J. Blackmon & C. H. Major (Eds.), *MOOCs and Higher Education: Implications for Institutional Research* (pp. 69–86). San Francisco, CA: Jossey-Bass.

Evans, B. J., Baker, R. B., & Dee, T. S. (2016b). Persistence patterns in Massive Open Online Courses (MOOCs). *The Journal of Higher Education, 87*(2), 206–242. doi:10.1080/00221546.2016.11777400

Gallie, D., & Paugam, S. (Eds.) (1999). *Welfare regimes and the experience of unemployment in Europe*. Oxford, UK: OUP Oxford.

Ganzeboom, H. B. G., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, *21*(1), 1–56. https://doi.org/10.1016/0049-089X(92)90017-B

Gerber, T. P., & Cheung, S. Y. (2008). Horizontal stratification in postsecondary education: Forms, explanations, and implications. *Annual Review of Sociology, 34*(1), 299–318. doi:10.1146/annurev.soc.34.040507.134604

Glass, C. R., Shiokawa-Baklan, M. S., & Saltarelli, A. J. (2016). Who takes MOOCs? *New Directions for Institutional Research, 2015*(167), 41–55. doi:10.1002/ir.20153

Gunn, C., McSporran, M., Macleod, H., & French, S. (2003). Dominant or different? Gender issues in computer supported learning. *Journal of Asynchronous Learning Networks, 7*(1). Retrieved from https://www.research.ed.ac.uk/portal/en/publications/dominant-or-different-gender-issues-in-computer-supported-learning(2afd96a8-45ec-4a9f-8b83-4df79245ec84).html

Ho, A., Chuang, I., Reich, J., Coleman, C., Whitehill, J., Northcutt, C., Williams, J., Hansen, J., Lopez, G., & Petersen, R. (2015). *HarvardX and MITx: Two years of open online courses Fall 2012-Summer 2014.* Retrieved from https://papers.ssrn.com/abstract=2586847

Hout, M. (2012). Social and economic returns to college education in the United States. *Annual Review of Sociology, 38*(1), 379–400. doi:10.1146/annurev.soc.012809.102503

Kizilcec, R. F., & Halawa, S. (2015). Attrition and achievement gaps in online learning. In *Proceedings of the Second ACM Conference on Learning @ Scale* (pp. 57–66). New York, NY: Association for Computing Machinery. doi:10.1145/2724660.2724680

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK'13) (pp. 170–179). New York, NY: Association for Computing Machinery. doi:10.1145/2460296.2460330

Kizilcec, R. F., & Saltarelli, A. J. (2019, June 24). Can a diversity statement increase diversity in MOOCs? In *Proceedings of the Sixth (2019) ACM Conference on Learning @ Scale* (pp. 1-8). New York, NY: Association for Computing Machinery. doi:10.1145/3330430.3333633

Kollock, P., & Smith, M. (1996). Managing the virtual commons: Cooperation and conflict in computer communities. In S. Herring (Ed.), *Computer-Mediated Communication: Linguistic, Social, and Cross-Cultural Perspectives* (pp. 109–128). Amsterdam, NL: John Benjamins. Retrieved from https://www.researchgate.net/publication/215643656_Managing_the_Virtual_Commons_Cooperation_and_Conflict_in_Computer_Communities

Leeper, T. J. (2018). *Interpreting regression results using average marginal effects with R's margins*. Retrieved from https://cran.r-project.org/web/packages/margins/vignettes/TechnicalDetails.pdf

Literat, I. (2015). Implications of massive open online courses for higher education: Mitigating or reifying educational inequities? *Higher Education Research & Development, 34*(6), 1164–1177. doi:10.1080/07294360.2015.1024624

Lucas, S. R. (2001). Effectively Maintained Inequality: Education Transitions, Track Mobility, and Social Background Effects. *The American Journal of Sociology*, *106*(6), 1642–1690.

Mare, R. D. (1981). Change and stability in educational stratification. *American Sociological Review, 46*(1), 72–87.

Merton, R. K. (1968). The Matthew effect in science. *Science, 159*(3810), 56–63.

OECD. (2018). *Youth not in employment, education or training (NEET)* [Data set]. OECD. doi:10.1787/72d1033a-en

Reich, J. (2014). MOOC completion and retention in the context of student intent. *Educause Review*. Retrieved from https://er.educause.edu/articles/2014/12/mooc-completion-and-retention-in-the-context-of-student-intent

Rocca, K. A. (2010). Student participation in the college classroom: An Extended multidisciplinary literature review. *Communication Education, 59*(2), 185–213. doi:10.1080/03634520903505936

Roksa, J., Grodsky, E., Arum, R., & Gamoran, A. (2007). United States: Changes in higher education and social stratification. In Y. Shavit, R. Arum, & A. Gamoran (Eds.*), Stratification in Higher Education: A Comparative Study* (pp. 167–191). Stanford, CA: Stanford University Press.

Rosenbaum, J. E., Deil-Amen, Regina, & Person, A. E. (2009). *After admission: From college access to college success*. New York, NY: Russell Sage Foundation.

Ruthotto, I., Kreth, Q., Stevens, J., Trively, C., & Melkers, J. (2020). Lurking and participation in the virtual classroom: The Effects of gender, race, and age among graduate students in computer science. *Computers & Education, 151*, 103854. doi:10.1016/j.compedu.2020.103854

Sallaz, J. J., & Zavisca, J. (2007). Bourdieu in American Sociology, 1980–2004. *Annual Review of Sociology, 33*(1), 21–41. doi:10.1146/annurev.soc.33.040406.131627

Schizzerotto, A., & Barone, C. (2012). *Sociologia dell'istruzione* [Sociology of education]. Bologna, IT: il Mulino

Shah, D. (2019). *Year of MOOC-based degrees: A Review of MOOC stats and trends in 2018 — Class Central. Class Central's MOOC Report*. Retrieved from https://www.class-central.com/report/moocs-stats-and-trends-2018/

Sharrock, G. (2015). Making sense of the MOOCs debate. *Journal of Higher Education Policy and Management, 37*(5), 597–609. doi:10.1080/1360080X.2015.1079399

Shavit, Y., Arum, R., Gamoran, A., & Menachem, G. (Eds.) (2007). *Stratification in higher education: A Comparative study*. Stanford, CA: Stanford University Press.

Shavit, Y., & Blossfeld, H.-P. (Eds.) (1993). *Persistent inequality: Changing educational attainment in thirteen countries*. Boulder, CO: Westview Press.

Triventi, M. (2013). The Role of higher education stratification in the reproduction of social inequality in the labor market. *Research in Social Stratification and Mobility, 32*, 45–63. doi:10.1016/j.rssm.2013.01.003

Triventi, M., Vergolini, L., & Zanini, N. (2017). Do individuals with high social background graduate from more rewarding fields of study? Changing patterns before and after the "Bologna process." *Research in Social Stratification and Mobility, 51*, 28–40. doi:10.1016/j.rssm.2017.07.001

van de Oudeweetering, K., & Agirdag, O. (2018a). MOOCS as accelerators of social mobility? A Systematic review. *Educational Technology & Society, 21*(1), 1–11.

van de Oudeweetering, K., & Agirdag, O. (2018b). Demographic data of MOOC learners: Can alternative survey deliveries improve current understandings? *Computers & Education, 122*, 169–178. doi:10.1016/j.compedu.2018.03.017

Wang, Y., & Baker, R. (2018). Grit and intention: Why do learners complete MOOCs? *The International Review of Research in Open and Distributed Learning, 19*(3). doi:10.19173/irrodl.v19i3.3393

Weber, M. [1922] 1978. *Economy and Society*. Berkeley, CA: University of California Press.

Zhang, D. J., Allon, G., & Van Mieghem, J. A. (2017). Does social interaction improve learning outcomes? Evidence from field experiments on massive open online courses. *Manufacturing & Service Operations Management, 19*(3), 347–367. doi:10.1287/msom.2016.0615

Appendix A

Statistical learning: Estimates from logistic regression (odds ratio) for both outcome variables

	Model 1	Model 2
	Outcome = certificate	Outcome = engagement
employment status (ref= FT employed)		
PT employed	1.019	1.074
	(0.109)	(0.115)
Student	1.107	1.141
	(0.0844)	(0.0874)
Unemployed	1.323**	1.196
	(0.126)	(0.117)
Retired	1.160	1.075
	(0.335)	(0.294)
Education (ref= less than BA)		
Bachelor's	0.945	0.986
	(0.109)	(0.117)
Doctorate	1.755**	1.734**
	(0.235)	(0.238)
Master's/professional degree	1.257	1.281*
	(0.150)	(0.157)
Withheld	0.942	1.333
	(0.247)	(0.347)
Sex		
Age		
Constant	0.214**	0.345**
	(0.0276)	(0.0457)
Observations	9,179	7,239

Note. Standard errors in parentheses; $^{*}p < .05$; $^{**}p < .01$.

Appendix B

America's Poverty & Inequality. Estimates from logistic regression (odds ratio) for both outcome variables

	Model 1 Outcome = certificate	Model 2 Outcome = Engagement
Employment status		
Unemployed	1.331	2.349*
	(0.446)	(0.872)
Student	0.934	1.008
	(0.285)	(0.311)
Retired	1.213	1.664
	(0.357)	(0.518)
Other	2.562**	1.965*
	(0.793)	(0.640)
Education		
Bachelor's	1.573	1.540
	(0.540)	(0.546)
Doctorate	1.387	1.277
	(0.536)	(0.506)
Master's/professional degree	1.173	1.304
	(0.401)	(0.457)
Withheld	0.939	0.710
	(0.735)	(0.590)
Parents' education		
Associate/some college	1.097	1.202
	(0.216)	(0.242)
Bachelor's	1.096	1.203
	(0.389)	(0.446)
Master's/phd	1.395	1.412
	(0.274)	(0.287)
Sex	Y	Y
Age	Y	Y
Constant	0.370*	0.536
	(0.150)	(0.220)
	0.0369	0.0449
Observations	824	711

Note. Standard errors in parentheses; $^{*}p < .05$; $^{**}p < .01$.

Chu, H.-C., Tsai, W. W.-J., Liao, M.-J., Chen, Y.-M., Chen, J.-Y. (2020). Supporting E-Learning with Emotion Regulation for Students with Autism Spectrum Disorder. *Educational Technology & Society*, *23* (4), 124–146.

Supporting E-Learning with Emotion Regulation for Students with Autism Spectrum Disorder

Hui-Chuan Chu[1], William Wei-Jen Tsai[2], Min-Ju Liao[3], Yuh-Min Chen[2*] and Jou-Yin Chen[4]

[1]Department of Special Education, National University of Tainan, Taiwan // [2] Institute of Manufacturing Information and Systems, National Cheng-Kung University, Taiwan // [3]Department of Psychology, National Chung-Cheng University, Taiwan // [4]Kaohsiung Municipal Hsin-Da Elementary School, Taiwan // huichu@mail.nutn.edu.tw // xturtle@imis.ncku.edu.tw // mjliao@ccu.edu.tw // ymchen@mail.ncku.edu.tw // chenzuin@gmail.com

[*]Corresponding author

(Submitted January 30, 2020; Revised March 22, 2020; Accepted October 6, 2020)

ABSTRACT: Students with Autism Spectrum Disorder (ASD) in general have been found to have significantly lower academic achievement relative to their level of ability. Research has shown that students' emotional impairment with ASD severely interferes with their learning process, and academic emotions are domain-specific in nature. Therefore, the regulation of domain-specific academic emotions is an important approach to help students with ASD learn effectively. This study proposed an e-learning model that featured emotion recognition and emotion regulation to enhance mathematics e-learning for students with ASD. An emotion recognition approach based on facial recognition, an emotion regulation model, and a mathematics e-learning platform, were developed to realize the e-learning model. Two e-learning conditions: timed contest and increased difficulty of learning, were created for gathering information by observing two indexes: mathematical learning performance and negative emotional behaviors in each condition. An experiment in a mathematical e-learning context was performed to evaluate the performance of e-learning and emotion regulation effectiveness. The results of the emotion recognition classifier reached a 93.34% average recognition rate, and the participants of this experiment displayed a statistically significant decrease in targeted negative behaviors from baseline to intervention (p = .000) and significant improvements in mathematics learning performance (p = .005); however, responses to emotion regulation interventions varied among the participants. Implications for research and practice are discussed.

Keywords: Autism spectrum disorder, Emotion regulation, Emotion recognition, E-Learning, Intelligent systems, Mathematics learning

1. Introduction

Educational reformation is a global trend in the 21st century, with an emphasis on improving the learning competence of students with disabilities. However, since many students with disabilities learn differently from other students, one of the emphases of special education is to provide adaptive education in order to develop the strengths of students based on their characteristics.

As the Internet and other computer technology continue to develop rapidly, e-learning has become an emerging learning method that can provide adaptive learning services, thereby significantly improving learning effectiveness. E-learning has also been employed in the field of special education, and has demonstrated learning improvements of students with autism spectrum disorder (ASD) (Khowaja & Salim, 2013). Research has shown that children with ASD may learn more rapidly when tasks are presented by a computer rather than by a teacher (Heimann, Nelson, Tjus, & Gillberg, 1995; Moore & Calvert, 2000). In addition, e-learning enables students with ASD to use their strengths in visual-spatial skills to overcome the challenges they encounter during learning (Cheng & Ye, 2010; Swezey, 2003).

Emotional impairment, one of the significant characteristics of students with ASD, can severely affect their daily activities, such as learning, communication, and interacting with teachers and peers. Totsika, Hastings, Emerson, Lancaster and Berridge (2011) found that 74% of children with ASD and no intellectual disability had clinically significant emotional difficulties, such as anger, sadness or anxiety, compared to 18% of typically developing peers. Other investigators have also reported that 17% to 74% of individuals with ASD exhibited anxiety-related behavior (Tsai, 1999). Intense anxiety can result in syndromes such as obsessive-compulsive disorder, generalized anxiety disorder, social anxiety, panic disorder, phobia and depression (Bejerot, 2007; Leyfer, Folstein, Bacalman, Davis, Dinh, & Morgan, 2006; MacNeil, Lopes, & Minnes, 2009). Also, student with ASD often exhibit impaired cognitive flexibility (Hughes, Russell, & Robbins, 1994; Kriete & Noelle, 2015) that makes students with ASD frustrated more than typical peers when facing new and difficult problems. Furthermore, conventional learning contexts are often time-limited, such as test (Roskam, 1997) and include increased difficulties of academic demands resulting in unexpected situation - the ordeal of learning makes students with ASD cannot accomplish their learning mission in the designated time period, which then induced negative academic emotions. Ashburner, Ziviani, and Rodger (2010) compared teacher ratings of academic performance and classroom emotional and behavioral regulation of 28 students with ASD (with average range IQ) and 51 typically developing students drawn from the same mainstream classrooms, and teachers rated students with ASD as exhibiting emotional and behavioral difficulties (including attention difficulties, anxiety, depression, oppositional and aggressive behaviors) to a significantly higher level than their typically developing peers. These negative academic emotions interfere with learning, resulting in unfavorable learning outcomes (Bejerot, 2007; MacNeil et al., 2009). As shown in the study of (Ashburner et al., 2010), fifty-four percent of students with ASD were rated by teachers as under-achieving academically as compared to 8% of typically developing students. However, domain-specific academic emotions in ASD has been understudied.

Mathematics problem solving is necessary for daily life (Lerner & Kline, 2006; Roman, 2004). However, in mathematics, one of the most difficult courses (Ginsburg, 1997), negative academic emotions appear very often in the learning process, especially for students with disabilities (Georgiou, Soulis, Rapti, & Papanikolaou, 2018) and greatly affect the learning outcomes (Carey, Hill, Devine, & Szücs, 2015). Oswald, Beck, Iosif, McCauley, Gilhooly, Matter, and Solomon (2016) found that the strongest predictor of math problem solving was perceptual reasoning, followed by verbal ability and test anxiety, then the diagnosis of ASD. Thus, reducing negative emotions of students with ASD during learning may be beneficial to enhancing their math abilities. Furthermore, mathematics problem solving is a process that combines reading, thinking, and computational skills (Forsten, 2004). If there is a methodology to benefit this process, by reducing their negative emotions, wider applications of other academic subjects would be anticipated. To sum up, factors related to mathematic learning could be categorized as individual and external. Individual factors include impaired emotion and cognition, whereas external factors could include difficulty and time limitations. The interactive effect of individual and external factors may induce negative emotions during the learning process. Therefore, assistance in regulating emotions is important for students with ASD to learn effectively in both authentic
and virtual learning environments.

Emotional assistance in the context of e-learning requires automatic emotion recognition and adaptive emotion regulation. Since the concept of affective computing was first proposed (Picard, 1997), e-learning platforms have evolved from intelligent tutoring systems (ITSs) (Schiaffino, Garcia, & Amandi, 2008; Curilem, Barbosa, & De Azevedo, 2007) to affective tutoring systems (ATSs) (Afzal & Robinson, 2011; Mao & Li, 2010; Shen, Wang, & Shen, 2009), in which emotion detection methods were developed to identify a student's emotional state automatically. Current methods for identifying emotional features include physiological signals, facial expressions, speech, and physical postures (Moridis & Economides, 2008). The facial-expression-based approach seems more appropriate when recognizing students' emotional states during the learning process for two reasons: First, human emotional expressions come in numerous forms, such as facial expressions, body language, and vocalizations. Among these, 55% of emotional information is conveyed by facial features (Mehrabian, 1968). Second, wearing a device for physiological sensors makes students feel uncomfortable and may cause them to be restive. Therefore, facial-expression-based recognition that does not need body-based devices would be more suitable in the learning process. However, as students with ASD have severe difficulties with communication-related emotions, their facial expressions change less than others (Bieberich & Morgan, 2004; Czapinski & Bryson, 2003).

Studies on emotional regulation have adopted various strategies and methods, including cognitive mediation strategy (Lehmkuhl, Storch, Bodfish, & Geffken, 2008; Singh, Lancioni, Singh, Winton, Singh, & Singh, 2011; Reaven, Blakeley-Smith, Leuthe, Moody, & Hepburn, 2012; Sze & Wood, 2007), attention transfer, learning context adjustment, and stepwise guides (Parkinson & Totterdell, 1999). However, previous investigations on emotion regulation for people with ASD have focused on regulating social emotions, not academic emotions. The effectiveness of emotion regulation depends heavily on the application of regulation strategies that are, in turn, based on the types of emotions, causes of negative emotions, and students' characteristics. Accordingly, a model of adaptive emotion regulation is required. The present study aimed to develop and validate an adaptive e-learning model featuring domain-specific academic emotion recognition and emotion regulation. For assessing the model's effectiveness, we attempted to answer the three other specific research questions:
(a) Are emotion regulation strategies effective in reducing negative academic emotions for students with ASD?
(b) Are emotion regulation strategies effective in enhancing learning performance for students with ASD?

(c) Are there differences in the effectiveness of emotion regulation strategies for students with mild (level 1) and moderate (level 2) ASD?

An e-learning model that featured emotion recognition and regulation to enhance e-learning for students with ASD was proposed in this study to answer these research questions. A facial expression-based emotion recognition approach, an emotion regulation model, and an e-learning platform were developed to realize the proposed e-learning model. To observe how interactive effects of individual and external factors influence learning emotion, this study manipulated external factors, including time limitation and increased difficulty of learning. Individual responses in e-learning were measured for mathematical learning performance and negative emotional behaviors in two conditions that were conducted in correspondence with the external factors: the timed contest and the increased difficulty of learning. To verify the feasibility and effectiveness of the proposed domain-specific academic emotion recognition and emotion regulation model, we conducted an experiment with 8 students with ASD. Considering the limitation of the small sample size, we aimed to provide some preliminary insights into the impact of this model on reducing negative academic emotions and improving the mathematical learning performance of students with ASD.

2. Literature review

2.1. E-Learning environment for students with ASD

E-Learning delivers a stable interactive environment that benefits not only learning but also social interaction. Various kinds of e-learning applications for ASD have been proposed, including academic learning, linguistic learning, and social interactions. This work surveyed the related literature, as follows: Colby (1973) utilized a composed voice to treat nonspeaking children with ASD and showed that 13 out of 17 children showed significant linguistic improvement. Cheng and Ye (2010) developed a collaborative virtual learning environment (CVLE-3D) to improve empathy instruction, which reached significant effectiveness. Golan and Baron-Cohen (2006) used mind-reading software to deliver emotion-related content to aid students with ASD by recognizing complex emotions. Similarly, Bernard-Opitz, Sriram, and Nakhoda-Sapuan (2001) developed an e-learning environment for instruction to help students with ASD deal with social conflict situations. Some studies used AR/VR techniques to benefit the learning process. Hu and Han (2019) used gesture-based instructions via VR technology to teach matching skills to age-schooled students with ASD. Politis, Sung, Goodman, and Leahy (2019) train people with ASD conversational skills through VR technology and demonstrates the potential of VR in participant feedback. Cakir and Korkmaz (2019) used AR-based learning materials for children with special educational needs and reached significant levels of improving learning performance, not only bringing them real-life experiences but also activating their motivation for learning.

However, the aforementioned researches only focused on improving social interactions and learning experiences; very few studies investigated methods for reducing students' academic negative emotions. Therefore, this research aimed to develop an e-learning model with emotion recognition and regulation to assist students with ASD in regulating negative emotions and improve their learning performance.

2.2. Emotion regulation

Current researches on emotion regulation for students with ASD have focused on emotions related to social interactions. Conner, White, Scahill, and Mazefsky (2020) evaluated the association between anxiety and emotion regulation from 1107 children with ASD and their parents; the result indicated emotion regulation impairment significantly predicted whether children's anxiety elevated. Konstantareas and Stewart (2006) investigated children with ASD regarding the relationship between affect regulation method and their temperament, by exposing them to a mildly frustrating situation and using the Children's Behavior Questionnaire (CBQ) to assess their temperament. The variance of the affect regulation results was more extensive than that of the controls, and the affect regulation method was thusly less effective. Farmer and Oliver (2005) assessed the social-emotional adjustment of children with pragmatic difficulties. The analytical results showed that children with ASD have more social-emotional adjustment difficulties than other children do. Sze's case study (Sze & Wood, 2007) described evidence-based cognitive-behavioral intervention and demonstrated successful treatment outcomes of an 11-years old girl with high-functioning autism (HFA). Lehmkuhl et al. (2008) demonstrated the effectiveness of the cognitive-behavioral intervention, successfully treating a 12-years-old male with ASD. Singh et al. (2011) used a mindfulness-based procedure for adolescents with Aspergers (a form of autism characterized by average or above-average intelligence, which eliminated from a subtype of ASD after DSM-V published) to control their aggressive behaviors, with no negative behavior observed near the end of the procedure. Reaven et al. (2012) recruited 24 adolescents with ASD and developed a program for the regulation of anxiety, in which nearly 46% of participants responded positively to the treatment. Conner, White, Beck, Golt, Smith, and Mazefsky (2019) proposed the emotional awareness and skills enhance (EASE) program, demonstrate significant clinical effectiveness and the feasibility.

Hilton, Ratcliff, Collins, Flanagan, and Hong (2019) examined a large data set of children from age 6 to 17 years and confirmed the need to address positive emotions and positive affect for individuals with ASD at a young age. The emotion regulation model proposed by Gross and Thompson (2007) delineated five aspects of the emotion regulation process: situ*ation selection, situation modification, attentional deployment, cognitive changes*, and *response modulation.* Quoidbach, Mikolajczak, and Gross (2015) further proposed that positive emotions can be increased through these 5 families of emotion regulation strategies. As noted in their recent review of studies which measured short-term increases in positive emotions, attentional deployment, cognitive change, and response modulation strategies have received the most empirical support (Quoidbach et al., 2015). These findings provide the basis for our research, and therefore the present study focused on developing these three types of the emotion regulation interventions that were modified and adapted to meet the needs of students of ASD.

2.3. Emotion regulation strategies

(a) *Situation modification* refers to tailoring a situation to modify its emotional impact. It is also referred to as problem-focus coping (Lazarus & Folkman, 1984) or primary control (Rothbaum, Weisz, & Snyder, 1982). To modify the situations, adapting teaching materials or curricula can be useful. McLaughlin (1993) claimed that most curriculum options of students with disabilities require adaptation to existing standard materials. Sometimes no adequate or easily adapted materials can be found; therefore, developing new curricula is needed. Furthermore, Hoover and Patton (1997) proposed that four perspectives should be considered when adapting the curriculum for students with disabilities: curriculum content, teaching strategies, teaching environment, and student behavior. When learning material is too hard, reducing the difficulty level or the amount of homework, simplifying steps, or giving tips are useful strategies. Van-Bockstaele, Atticciati, and Hiekkaranta (2020) applied situation modification in a negative situation to regulate their emotions by changing the environment.

(b) *Cognitive change* refers to selecting from the many possible meanings that may be attached to that situation. It is often used to either increase or decrease the emotional response or even change the emotion itself (Gross, 2002; Sze & Wood, 2007; Lehmkuhl et al., 2008; Chorpita & Daleiden, 2009; Reaven, 2010). Consequently, students' negative behavior can be reduced when they face difficult situations utilizing cognitive behavior therapy (CBT). Some studies have suggested that CBT, such as graded exposure and cognitive restructuring, effectively reduces anxiety in teenagers and children with ASD (Chorpita et al., 2009; Reaven, 2010; Sze et al., 2007). Maughan and Weiss (2017) pointed out CBT improved children's depression and positively affected parents across all parents in depression, emotion regulation, perceptions of their children, and mindful parenting when they joined the treatment process.

(c) *Attentional deployment* refers to how individuals direct their attention within a given situation to regulate their emotions (Gross et al., 2007). Grolnick, Bridges, and Connell (1996) pointed out that individuals use various methods to balance their emotions when facing an emotionally arousing situation. Among these, shifting attention from the arousing stimuli is useful for children, including their visual and motor exploration and passive use of objects and active engagement with substitute objects. Hamilton (2000) indicated that shiny objects could easily attract participants with ASD.

(d) *Response modulation* refers to efforts made to alter physiological, experiential, or behavioral responses in a situation (Gross et al., 2007). This family involves the modulation of the bodily component of the emotion by acting directly on the body itself (Gross, 1998). Strategies belonging to the response modulation family are muscle relaxation, deep breathing and positive imagery (Chorpita & Daleiden, 2009; Reaven, Blakeley-Smith, Nichols, Dasari, Flanigan, & Hepburn, 2009). Floress, Zoder-Martell, and Schaub (2017) combined relaxation training with social skills training and reinforcement principles to effectively increase the frequency of the targeted social skills of a student with ASD, and the effects were maintained for up to 17 weeks after the training program ended.

2.4. Classifier of emotion

To understand the emotions of students during the learning process, classifying emotion recognition is necessary. Hence, this work surveyed related recent studies on the recognition of emotions, as summarized below.

Human emotional expressions adopt numerous forms, such as facial expressions, body language, and voice. Among these, 55% of emotional information is conveyed through facial features, inferred by changes in the eyes, eyebrows, nose, and mouth (Mehrabian, 1968). Therefore, changes in facial expression are critical indicators for judging emotion.

Kapoor, Burleson, and Picard (2007) used a template-based facial landmark tracker, posture sensors, skin conductance, and mouse behavior model to detect gaming frustration, achieving 79% accuracy. Tariq, Li, Zhou, Wang, Huang, Lv, and Han (2012) extracted facial features by incorporating Hierarchical Gaussianization (HG), Scale Invariant Feature Transform (SIFT), and Optic Flow (OF), with SVMs to recognize emotions and person identification, achieving 80% accuracy. Chen and Lee (2011) used EEG (Electroencephalograph, an electrophysiological monitoring method to record the electrical activity of the brain) with SVM to detect anxiety in a web-based one-to-one language learning environment, and achieved 79.71% accuracy.

However, very few researchers have investigated emotion recognition applications in e-learning environments for students with ASD, whose facial expressions change less than those of other people (Bieberich et al., 2004; Czapinski et al., 2003). Therefore, this research aimed to develop an emotion recognition classifier specifically for students with ASD in e-learning applications, capable of recognizing happy, calm, and negative emotions, including anger and anxiety, of understanding students' mood changes during the learning process and to provide effective regulation.

3. Proposed model and methods

3.1. E-Learning model with emotion regulation

The proposed e-learning model includes pre-testing, learning path planning, adaptive learning material selection, adaptive learning, and post-testing, as shown in Figure 1. Before engaging in the learning for the first time, students are required to participate in a pre-test, including a computerized adaptive test and learning and thinking style assessments.

After the pre-test, a tentative learning path is planned based on math concepts in the student's knowledge model. Learning materials are selected according to the math concepts and mathematics error preventive strategies, and are adapted to suit the diverse traits of students based on their strengths, weaknesses, learning performances, and thinking styles. An error preventive strategy is applied to support preventive teaching by predicting mathematical errors that students may make according to their strengths and weaknesses. The purpose of this strategy is to reduce the difficulty of problems that students encounter d learning and to enhance their motivation to learn. After each stage of the adaptive learning process, remedial learning may be required according to the types of errors the students make, and the learning path will be adjusted based on students' learning status. However, the goal of this study is to support e-learning with emotion regulation for students with ASD. Therefore, to focus on the effectiveness of emotion regulation, learning style, thinking style, and learning path planning is not conducted in an experimental context.

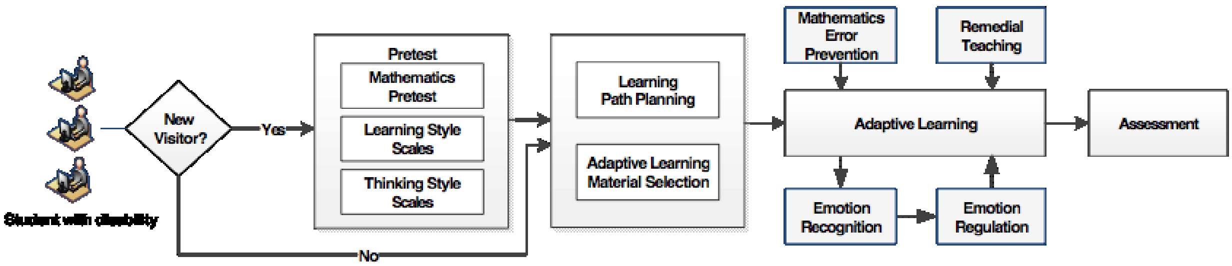


Figure 1. The proposed e-Learning model

During learning, an adaptive assessment is conducted to identify the students' learning progress and the types of errors made. This information is used to identify suitable teaching strategies for addressing these errors. Simultaneously, the affective states of students are monitored, and suitable regulation strategies are applied once unfavorable emotions are detected based on students' facial expressions. When students develop unfavorable emotions during the learning process, and regulation strategies cannot reduce these emotions, teachers can use remote support to intervene, helping students relax. After the entire learning process, the students take a computerized adaptive math test for the learning progress assessment. The next section introduces the main component of the proposed model, emotion regulation.

3.2. Emotion regulation

In this section, the categories and contents of the regulating strategies, the methods used in the decision-making for strategy selection, and the model of emotional regulation strategies are explained.

3.2.1. Emotion regulation strategies

The proposed regulation strategies can be categorized into three types: *response modulation, cognitive change, and attention deployment*.

Response modulation strategy: In this study, response modulation strategy is applied for regulating negative emotional behaviors, which involve a series of actions that will influence physiological, experiential, or behavioral responding as directly as possible (Gross et al., 2007). Strategies belonging to the response modulation family are muscle relaxation, deep breathing and positive imagery (Chorpita & Daleiden, 2009; Reaven et al., 2009). The computerized intervention sessions of the response modulation strategy were conducted with written and visual stimuli such as pictorial and graphical representations of the specific strategy steps that were presented in Figure 2. These provide students with ASD instruction on the implementation of each technique. Figure 2 illustrates an example of the computer sequence presented to a participant during the process of the response modulation strategy instruction.

Cognitive change strategy: In this work, the cognitive change strategies include social stories, self-instruction, and self-management. Appropriate strategies are selected for students to manage emotions and behaviors according to their personality traits and learning context. Each strategy involves a procedure, as well as voice and video clips. The procedure contains stepwise descriptions of the content of the cognitive intervention strategy. In each step, texts and video clips are used as tools for regulation. Students are introduced to the steps they should take to regulate their emotions and behaviors when encountering challenges. The strategy's design is inspired by Gross (2002), Sze and Wood (2007), Lehmkuhl et al. (2008), Chorpita & Daleiden (2009) and Reaven (2010). Figure 3 presents the steps that are involved in applying the social story strategy in a timed competition. The e-learning environment simulates the contexts presented in Figure 3 in the following order: descriptive sentences, perspective sentences, directive sentences, affirmative sentences, control sentences, or cooperative sentences. Integrated with a voice system, the model enables students to understand the meanings implied in the contexts and exhibit positive behaviors.

Attentional deployment strategy: This strategy helps directing the individual's attention towards positive aspects of the situation whereas focusing attention away from negative aspects (Campos, Frankel, & Camras, 2004; Gross, 1998). Positive distraction as an attentional deployment strategy, involves distracting oneself from negative emotions by engaging in activities that induce positive emotions to relax attention (Hanin, Grégoire, Mikolajczak, Fantini-Hauwel, & Van Nieuwenhoven, 2017). In this study, positive distraction included presentation of educational computer games, graphic animations, and task interspersal (Rapp & Gunby, 2016). The left image in Figure 4 presents an application of the game strategy. In this computer game, the students are instructed to draw on a whiteboard. The image on the right shows an animated geometry shape, and all animated geometry shapes (i.e., squares, rectangles, triangles, trapezoids, and polygons) were presented in a sequence order. Each shape was shown for 10 seconds, and the students with ASD were instructed to count the number of the geometry shapes that appeared during the intervention process.

The media presentation for the emotional regulation interventions was specifically designed for students with ASD and included texts, images, video clips, games, and graphic animations, which attracted learners' attention and enabled students with ASD to use their strengths in visual-spatial skills.

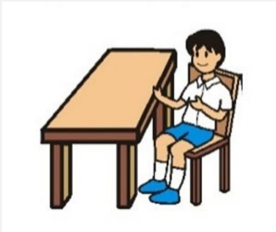
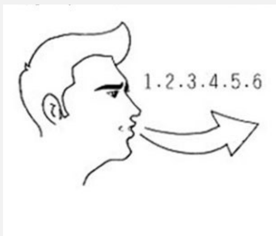
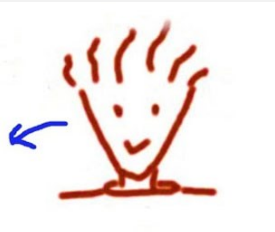
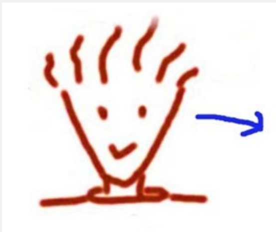
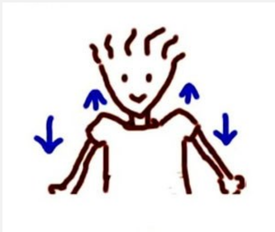
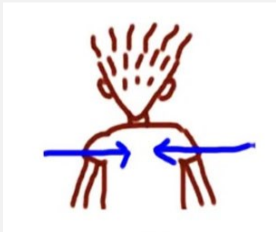
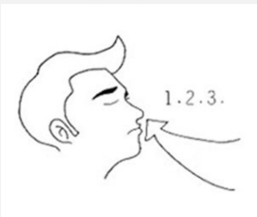
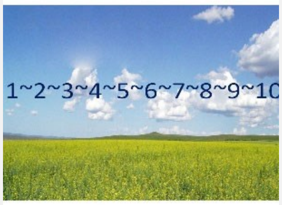
①	②	③	④
 When you are angry, you can follow these steps:	 Make yourself comfortable on a chair.	 Inhale and count to three.	 Exhale and count to six
⑤	⑥	⑦	⑧
 Make a fist with both hands and count to five.	 Relax both hands and count to five.	 Look up and count to three.	 Look down and count to three.
⑨	⑩	⑪	⑫
 Look to the left and count to three.	 Look to the right and count to three.	 Raise your shoulders toward your ears and count to three.	 Stretch your shoulders toward your back and count to five before you relax.
⑬	⑭	⑮	⑯
 Inhale and count to three.	 Lift both hands and cross your fingers over your head. Stretch your arms above your head and count to five.	 Make a fist with your left hand and use the fist to hit your right shoulder gently five times.	 Look at the picture of grass and imagine yourself lying on the grass. Count to ten.

Figure 2. Example of response modulation strategy

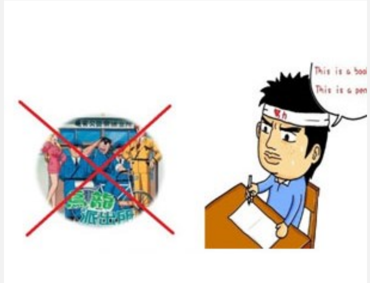
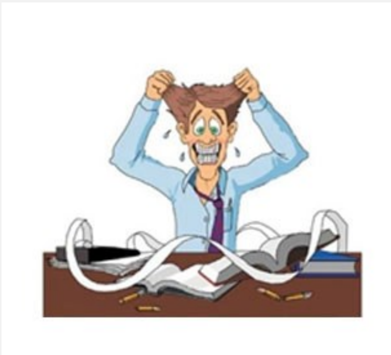
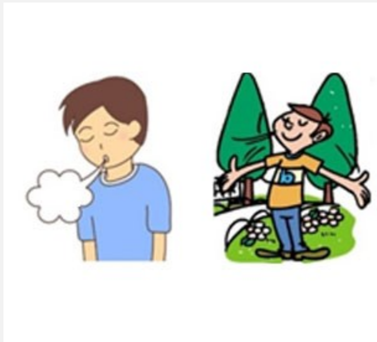
① descriptive sentences	② descriptive sentences	③ perspective sentences
 I love solving mathematical problems on the computer.	 Sometimes the homework has many questions like a test. I may not be able to finish all of them, but I must try my best.	 My mom always tells me that I must finish my homework before I can do things that I enjoy.
④ descriptive sentences	⑤ perspective sentences	⑥ directive sentences
 I realize that there are only two or three minutes left, but I still have many questions to finish.	 My mom and the teacher will worry if they see me panic.	 I can take three deep breaths and count from one to three while telling myself: "Take your time."
⑦ directive sentences		
 I can solve the problems one at a time.		

Figure 3. Example of cognitive change strategy: A social story

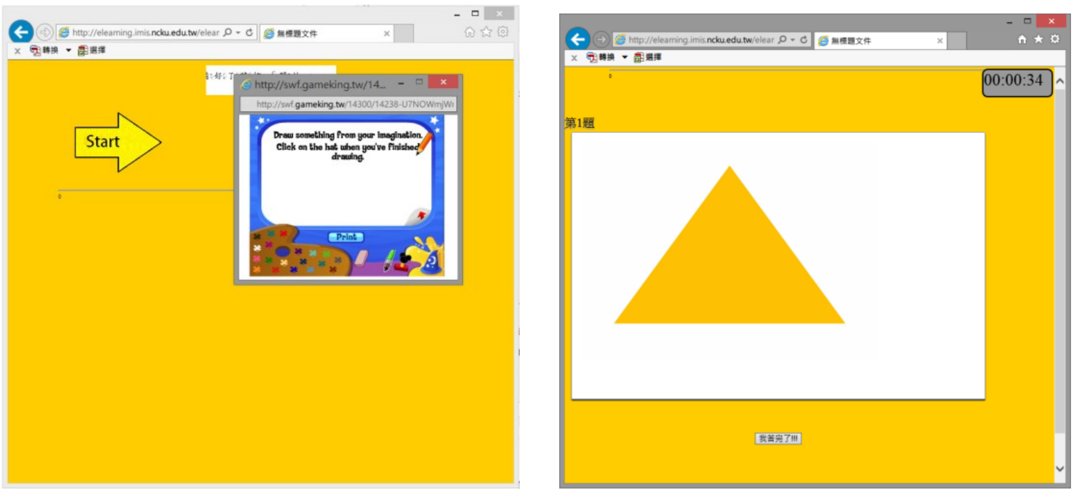


Figure 4. Example of attentional deployment strategy: Educational computer game and graphic animation

3.2.2. Model of strategy for regulating emotions

A preliminary model of emotional regulation strategies was developed based on the relations among personal background, types of learning emotions, and regulation strategies, through the following analysis: (a) the types of academic emotions that are exhibited by students with ASD, (b) the relationship between the personal background of students and their most frequent academic emotions, i.e., happiness, anxiety, anger and feeling calm, (c) the relationship between learning stress and academic emotions (Carey et al., 2015; Georgiou et al., 2018), and (d) corresponding regulation strategies.

The Fuzzy Delphi method (Dalkey & Helmer, 1963), an expert evaluation method, was employed to evaluate the preliminary model. The purpose of this step was to obtain expert opinions for the revision of the preliminary model. Based on the preliminary model, a fuzzy Delphi questionnaire was designed. Scholars and expert teachers in the ASD field were invited to complete the questionnaire by weighting the relations between types of academic emotions and corresponding strategies specified in the preliminary model. The relative significance of regulation strategies was calculated and defined. The simple center of gravity method was used to calculate the threshold values. All of the most agreed upon variables were incorporated into a Bayesian network analysis that produced a model for adaptive academic emotion regulation for students with ASD. This model was upgraded and dynamically maintained to ensure effectiveness of emotion regulation interventions by incorporating the participant data and the experts' opinion into the model during the application stage through a Bayesian network analysis (Xenos, 2004).

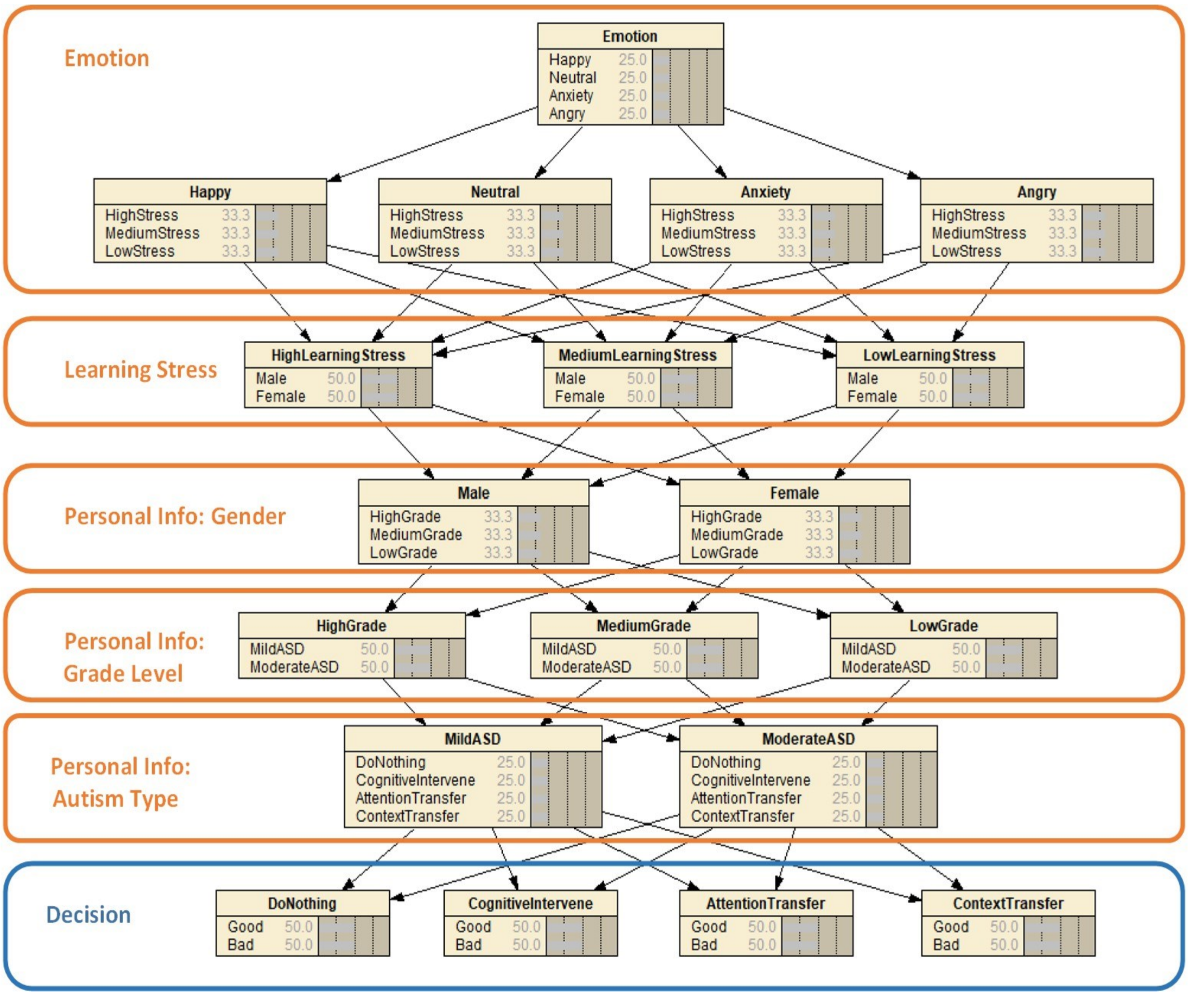


Figure 5. Structure of the Bayesian model for making decisions about choosing emotional regulation strategies

The Bayesian-based emotion regulation model has the following levels: emotion, learning stress, personal information (gender, grade, and ASD type), and the decision of strategy selection, as shown in Figure 5. The emotional state is captured from an emotion recognition classifier in real-time. Learning stress is predefined during the learning progress, which is decided by the difficulties of the current learning materials, and personal information is collected from the results of students' pretest and psychologists' diagnosis. The decision involves choosing one of the three aforementioned types of strategy from the repository of emotion regulation strategies.

3.2.3. Method for emotion regulation strategy selection

According to the principle of adaptability, choices of regulation strategies should be based on students' personal information and their specific emotional context. Numerous decision methods are accurate only when the sample is large enough. However, obtaining large amounts of above-mentioned information is often difficult and involves privacy concerns. Small samples suffice when Bayesian networks are employed to make decisions; furthermore, training, and reflections take little time (Mendes & Mosley, 2008). This method satisfies the need to use real-time emotion regulation strategies to provide rapid responses; therefore, the Bayesian network was employed to determine appropriate emotion regulation strategies.

A Bayesian network is based on a directed acyclic graph (DAG) G = (I, E), where I denotes the combination of all nodes, and E denotes the combination of all edges. Assume that $X = (X_i)_{i \in I}$ is a random variable of node *i*. If the joint probabilistic distribution of X is obtained using the following equation, then X is the Bayesian network based on graph G, where *pa(i)* is the prior probability of node *i*:

$$p(x) = \prod_{i \in I} p(x_i | x_{pa(i)}) \qquad - (4)$$

After the learning process is completed, the outcome is assessed to evaluate the effectiveness of the selected regulation strategy. The results are fed back to the Bayesian network model, where the probability values of the nodes are modified to yield perfected regulation criteria.

3.2.4. The classifier for academic emotion recognition

To understand the academic emotions of students with ASD during the mathematics e-learning process, this work developed a classifier for recognizing emotion. The details are described below.

Two stages were involved (see Figure 6) in developing the method for facial expression-based emotion recognition: *facial feature extraction* and *recognition model construction* (i.e., the emotion classifier). An emotion elicitation experiment was performed in a mathematical e-learning environment to collect facial-based features for training the emotion classifier. The participants of this experiment were fifteen students (14 boys) with ASD (aged 8 to 12 years) recruited from elementary schools. All participants were diagnosed with ASD by psychiatrists based on the current Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-V) diagnostic criteria (American Psychiatric Association, 2013).

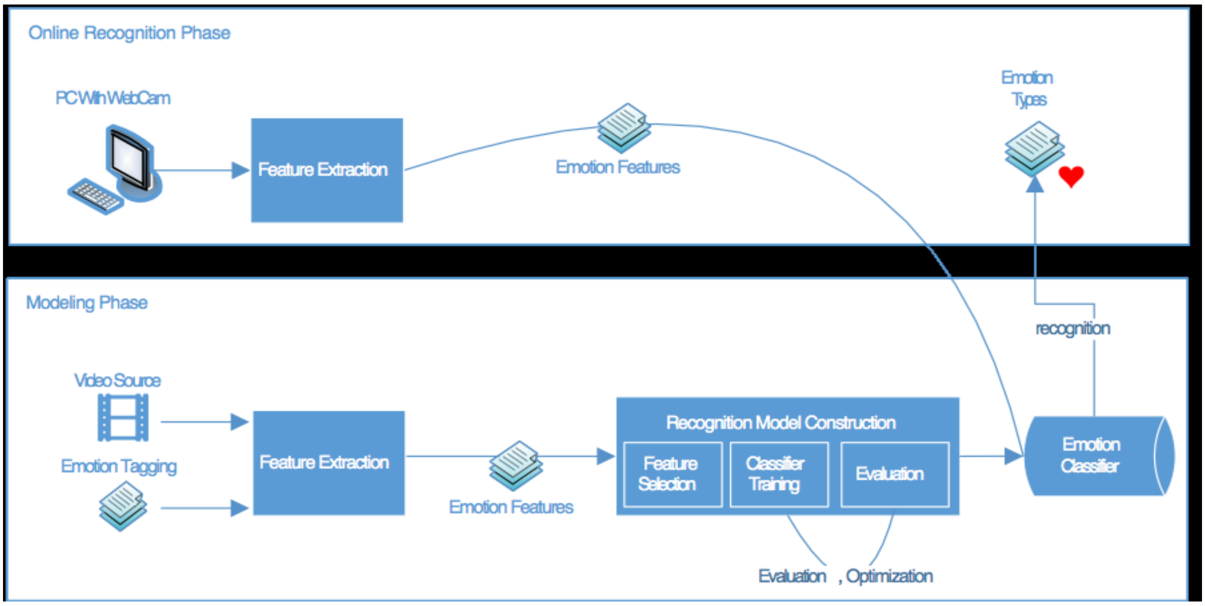


Figure 6. Emotion recognition mechanism

In the stage of feature extraction, the Face Tracking API 3.2 system was used to track facial points and calculate facial features. The obtained facial signals were combined with emotion assessments made by expert teachers and the parents of the participants regarding the participants' emotional states, and 39,067 samples were collected from the experiment.

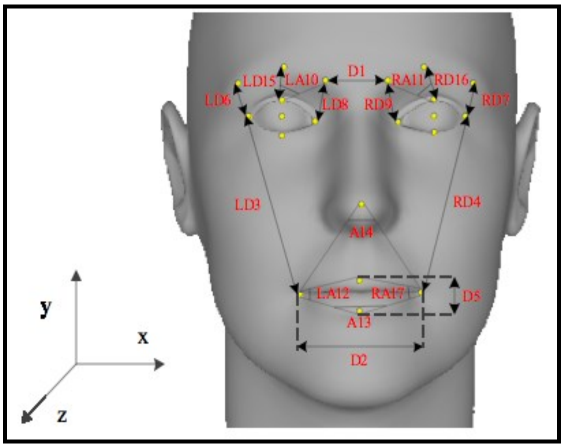
The support vector machine (SVM) (Cortes & Vapnik, 1995) was used to construct the classifier for emotion recognition. To determine robust features for emotion recognition, Information Gain (IG) (Quinlan, 1979) and Chi-Square were used for feature evaluations. The effectiveness of the classifiers with different parameters of sliding windows was also examined.

Facial Feature Extraction was used to generate feature vectors based on the facial landmark coordinate signals extracted by the Face Tracking API. This process has three steps: coordinate conversion, statistical processing, and normalization. The generated feature vectors can be used as the training samples of the classifiers. Figure 7 shows the distribution of facial features.

Coordinate conversion: Distance and angle were used as the two observable metrics to identify facial expression variations. All facial coordinates were transformed to distances and angles, yielding 12 distances and five angles, as shown in Figure 3. Equations (1) and (2) use distance D5 and angle A14 as examples.

$$D5 = \sqrt{(mP2x - mP4x)^2 + (mP2y - mP4y)^2} \quad \text{- (1)}$$

$$A14 = cos^{-1} \frac{(mP1nP)\cdot(mP3nP)}{|mP1nP|\cdot|mP3nP|} \quad \text{- (2)}$$



Where:

D5: the distance between the top and bottom of the opened mouth.
A14: the angle between both corners of the mouth is subtended at the nose.
mP1: right corner of the mouth.
mP2: center point of the upper lip.
mP3: left corner of the mouth.
mP4: center point of the bottom lip.
nP: nose.

Figure 7. Facial feature distribution

Statistical processing: To identify the trends in the signal variations, the statistics concerning the observed values were calculated; these were: maximum and minimum values, standard deviations, and means. A total of 34 features [(12 distance values + 5 angle values) × 2 statistics] were generated, yielding 34 × 1 feature vectors.

Normalization: When the range of values of a feature is relatively wide, the feature may influence the operation of the classifier. Hence, all features' ranges of values must be normalized to ensure that the features contribute proportionally. Equation (3) was used for normalization.

$$x' = \frac{x - min}{max - min} \quad \text{- (3)}$$

where

x': normalized value. x: the value of original signals. min: the minimum value of the original signal. max: the maximum value of the original signal.

Finally, the generated feature vectors are used as training samples to construct a classifier. Each sample represents the variations of the facial features and expressions exhibited by each participant within 30 sec. A sample consists of a set of feature vectors and an emotion label. The use of the samples to construct the classifier and identify useful feature combinations is explained below.

The development of the recognition model includes three steps: *feature selection, classifier training*, and *evaluation*, as shown in Figure 2. Before *classifier training*, *feature selection* was conducted to assess the features that facilitate the construction of a highly accurate classifier and to reduce the dimensionality of the input feature vectors. Thus, the performance of the classifier could be improved. To meet the requirements of practical applications, real-time recognition of negative emotions is necessary. Therefore, feature selection should reduce the dimensionality of the input vectors. Consequently, attention can be focused on critical features, reducing the response time and the classifier's construct time.

The *feature selection* algorithms used in this model included filters and wrappers. Filter algorithms evaluate all features individually, using threshold values to remove features that do not contribute enough to the classifier (Guyon & Elisseeff, 2003). Wrapper algorithms identify optimal feature combinations iteratively, using classification algorithms to evaluate various selected feature combinations (Guyon et al., 2003). Since filter algorithms take relatively little time to execute, they were used to identify the optimal feature combinations. Among the filter algorithms, Information Gain and Chi-Square methods were used to evaluate the contribution of features, and the top 50% of features were used for classifier construction.

An SVM was employed to construct the classifier by iteratively feeding each feature vector from the training set, including 39,150 facial samples. Parameters of the SVM classifier (cost C and gamma γ) were optimized by the grid search method (Huang, Chen, & Wang, 2007). We found the best (C, γ) combination to be (C = 16.0, γ = 10.0) in this work. The feature vectors generated in the feature extraction stage were used as the training set. To verify the effectiveness of the classifiers, 10-fold cross-validation was used to avoid over-fitting (Kohavi, 1995). The Precision/Recall index was used in the evaluation protocol.

In the application of e-Learning, the emotion classifier records 30 sec of facial feature variations that are exhibited by the students before converting the signals into vectors that can be used as input to the emotion recognition model. The emotions that were exhibited by the students were thus determined.

4. Experiment

This section presents the experimental design to verify the effectiveness of the proposed academic emotion recognition and regulation based mathematics e-learning model for students with ASD.

4.1. Experimental design

This was an intervention study design consisting of baseline and alternating treatments conditions, and the dependent variables of this study were the percentages of negative emotional behaviors and the rates of mathematical learning performance. In the baseline phase, emotional behaviors in computer simulation environments were initially observed in two e-learning conditions: *timed contest* and *increased difficulty of learning*. Data were collected regarding the emotions of each participant in response to each computer simulation environment. Emotions were automatically detected based on facial features (as explained in Section 3.3) during the alternating treatments conditions, which were counterbalanced to control for sequence effects. Emotion regulation strategies were used on two occasions to assist in the learning process. The first occasion was when negative emotion was detected, and the proposed Bayesian model would suggest an adaptive strategy. The other was when the teacher determined that negative emotional behavior had appeared. Finally, using observational data concerning negative emotional behaviors, the effectiveness of the emotion regulation strategies was evaluated. Data on the rates of mathematical learning performance were also collected to analyze whether reducing negative emotional behaviors may improve learning performance across the two e-learning conditions. Table 1 presents the defined negative emotional behaviors in this research (Pierce & Courchesne, 2001; Patnam, George, George, & Verma, 2017; Rivard & Forget, 2012).

Table 1. Negative emotional behaviors

Type	Presentation
Stereotyped Movement (Pierce et al., 2001)	Clicking mouse frequently, scratching, and biting fingers, shaking a leg
Verbal Behavior (Rivard et al., 2012)	Complaining, calling parents, or clucking
Facial Expression (Patnam et al., 2017)	Crying, with tears in eyes

The experiment was conducted in a bright and quiet room. Parents, a teacher, and two researchers observed participants via monitor outside the room. The descriptions of the duration and activities of each e-learning condition of the experiment are shown in Table 2. The duration of emotion regulation across two e-learning conditions was three minutes. If emotion regulation happened in the alternating treatment period, time would be extended according to how long the emotion regulation intervened. The answer time in the timed contest condition was five minutes for each question. A set of questions that slightly exceed the participants' ability would be delivered in the difficulty learning condition. When the terminating condition was satisfied, the stage would be finished. Besides this, three-minute breaks were taken in-between the two conditions in the alternating treatment period. During the break, participants were provided with drinks or snacks of their choice, allowing them to take a short rest. The process was conducted at the same time each day.

Table 2. The descriptions of the duration and activities of each e-learning condition of the experiment

Activities	Timed Contest (TC)	Difficulty Learning (DL)
Emotion regulation Intervention	3min	3min
Time limitation for solving the problems	5min	N/A
Breaks between conditions	3min	3min
The criteria for discontinuing the session	Answered 4 questions	Duration more than 20min

4.2. Participants

Eight students with ASD aged 8 to 14 (M = 9.5 years, SD = 1.92), consisting of seven boys and one girl, participated in this study. Seven participants were enrolled in special education resource rooms at an elementary school, and the remaining participant was enrolled in a special education resource room at a middle school. All participants were diagnosed with ASD by psychiatrists based on the current DSM-V diagnostic criteria. The participants attended regular classrooms for more than half of each day at school while being mentored by teachers who specialized in special education. Participants' prerequisite abilities were (a) a requirement to have basic mathematical knowledge, language, and communication skills. (b) the ability to operate the e-learning system interface using a keyboard and mouse. And (c) were evaluated using the Wechsler Intelligence Scale for Children-Third Edition (WISC-III), in which verbal or performance IQ at or above 70. All participants satisfied the above criteria. Table 3 presents the demographic information on the participants.

Table 3. Participant characteristics

	Gender	Chronological age	ASD severity
1	F	10	Moderate (Level 2)
2	M	9	Mild (Level 1)
3	M	14	Mild (Level 1)
4	M	8	Mild (Level 1)
5	M	9	Mild (Level 1)
6	M	9	Moderate (Level 2)
7	M	9	Mild (Level 1)
8	M	8	Mild (Level 1)

4.3. Method of quantitative analysis

This study observed two quantitative indicators: the percentages of negative emotional behaviors and the rates of mathematical learning performance. Negative emotional behaviors (Table 1) were evaluated with 10s-partial-interval recording, and the target behavior was recorded if the participant engaged in the behavior at any time during a given 10-s interval. Data were then converted to a percentage of occurrences of negative emotional behaviors for a given session. The mathematical learning performance rates were defined as the number of math problems the participant completed correctly per minute. A small-sample analysis (n = 8) was performed. As the Wilcoxon signed-rank test has been proven to be representative for small sample sizes (Sidney, 1956), it was used herein to evaluate the effectiveness of changes, and we assessed effect sizes by calculating d-statistics for pairwise comparisons (Dunlap, Cortina, Vaslow, & Burke, 1996).

5. Data analysis and results

5.1. Results of emotion recognition classifier

Table 4 shows the performance of emotion recognition. By using the SVM classifier and Information Gain of the feature selection, the average overall recognition rate was 93.34%. Additionally, to prove the effectiveness of the proposed emotion recognition method, this research used ROC (Receiver Operating Characteristic), AUC (Area Under Curve), and a standard technique for summarizing classifier performance over a range of trades using true positive and false positive error rates (Swets, 1988). The ROC curve is a robust method of identifying potentially optimal classifiers (Provost & Fawcett, 2001). When ROC AUC is over 0.7, the classifier has enough discernibility to recognize emotion. In Table 4, all the ROC AUC exceeded 0.7.

Table 4. Recognition rate

Emotion	Recognition rate	ROC AUC	F-measure
Calm	95.10%	.947	.921
Happy	88.40%	.939	.914
Anxiety	97.20%	.952	.936
Anger	85.70%	.927	.900
Overall	93.34%	.946	.924

These results displayed enough capability in recognizing four types of emotion and verified that the proposed e-learning model could deliver emotion regulation materials triggered in-time by emotion recognition with high accuracy.

5.2. Overall effects of the academic emotion regulation interventions for students with ASD

Table 5 shows the proportion of suggested regulation strategies by the Bayesian model. With regards to the overall changes in negative emotional behaviors and mathematical learning performance rates over time (see Table 6), Wilcoxon signed rank tests showed significant improvements in negative emotional behaviors (overall p = .000***) corresponding to large effect size (ES = 1.010). Mathematical learning performance rates significantly increased (overall p = .005*) with a moderate effect size (ES = .594) between baseline and intervention phases.

Table 5. The proportion of suggested regulation strategies by Bayesian model

Strategy type	Timed contest (TC)	Difficulty learning (DL)
Response modulation	58.8%	55.5%
Cognitive change	35.2%	38.9%
Attention deployment	5.8%	5.5%

Table 6. Changes between baseline and intervention phases

Dependent variable	Phases						p-value (ES)
	Baseline			Intervention			
	Median	Mean	Range	Median	Mean	Range	
NEB	53.00%	46.40%	8.42%; 90.83%	18.04%	21.86%	4.62%; 49.40%	.000*** (1.010)
MLPR	1.32	1.65	0.17; 5.22	1.66	2.51	0.28; 6.93	.005* (0.594)

Note. $^{*}p$ < .05; $^{***}p$ < .001. NEB: Negative Emotional Behavior, MLPR: Mathematical Learning Performance Rate.

5.2.1. Effects of interventions on negative emotional behaviors of students with ASD

Tables 7, 8 and 9 show change in negative emotional behaviors between baseline and intervention phases for both conditions (TC, DL), both levels of ASD (level 1, level 2) and the three different intervention strategies (RM, CC, AD). Participants demonstrated a statistically significant decrease in their negative emotional behaviors across both conditions (TC p = .004, DL p = .004) with large effect sizes (TC ES = .888, DL ES = 1.155). In addition, participants with mild (level 1) ASD showed a significant decrease during the intervention period in their negative emotional behaviors with a large effect size (p = .000, ES = .854). A moderate decrease was also observed in the participants with moderate (level 2) ASD; however, it was not statistically significant (p

= .063). This decrease was also significant for each of the strategy as shown in Table 9 (RM p = .000, CC p = .000, AD p = .004) with large effect sizes (RM ES = .987, CC ES = 1.011, AD ES = 2.720).

Table 7. Change in negative emotional behaviors between baseline and intervention phases for both conditions

Condition	Phases						p-value (ES)
	Baseline			Intervention			
	Median	Mean	Range	Median	Mean	Range	
TC	46.34%	41.66%	8.42%; 71.83%	18.04%	22.70%	4.62%; 49.40%	.004* (0.888)
DL	58.84%	51.15%	9.67%; 90.83%	17.74%	21.02%	6.53%; 46.08%	.004* (1.155)

Note. $^{*}p < .05$. TC: Timed Contest Condition, DL: Difficulty Learning Condition.

Table 8. Change in negative emotional behaviors between baseline and intervention phases for each level of ASD

Level of ASD	Phases						p-value (ES)
	Baseline			Intervention			
	Median	Mean	Range	Median	Mean	Range	
Moderate	65.92%	65.67%	40.00%; 90.83%	25.17%	25.26%	4.62%; 46.08%	.063 (2.190)
Mild	52.17%	39.38%	8.42%; 65.33%	18.04%	20.73%	6.53%; 49.40%	.000** (0.854)

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

Table 9. Change in negative emotional behaviors between baseline and intervention phases for the different strategies

Strategy	Phases						p-value (ES)
	Baseline			Intervention			
	Median	Mean	Range	Median	Mean	Range	
RM	53.00%	46.40%	8.42%; 90.83%	22.68%	22.40%	4.95%; 49.40%	.000*** (0.987)
CC	53.00%	45.74%	8.42%; 90.83%	14.68%	20.77%	3.92%; 49.85%	.000*** (1.011)
AD	61.52%	55.94%	30.33%; 71.83%	21.80%	20.16%	3.30%; 36.10%	.004* (2.720)

Note. $^{*}p < .05$; $^{***}p < .001$. RM: Response Modulation, CC: Cognitive Change, AD: Attention Deployment.

5.2.2. Effects of interventions on mathematical learning performance rates of students with ASD

Tables 10, 11 and 12 show change in mathematical learning performance rates between baseline and intervention phases for both conditions (TC, DL), both levels of ASD (level 1, level 2) and the three different intervention strategies (RM, CC, AD). As reported in Table 10, the Wilcoxon sign-ranked test approached significant increase (p = .074). for the number of math problems completed correctly per minute between baseline and intervention in the time contest condition (M = 2.28 vs. M = 2.79) with a moderate effect size (ES = .349). The number of math problems completed correctly per minute between baseline and intervention in the difficulty learning condition showed a similar tendency increasing from the baseline level of 1.01 per minute to the intervention level of 2.24 per minute, with a large effect size (p = .071, ES = 1.069).

Table 10. Change in mathematical learning performance rates between baseline and intervention phases for both conditions

Condition	Phases						p-value (ES)
	Baseline			Intervention			
	Median	Mean	Range	Median	Mean	Range	
TC	1.86	2.28	0.53; 5.22	2.18	2.79	0.41; 6.75	.074 (0.349)
DL	0.52	1.01	0.17; 3.67	1.00	2.24	0.28; 6.93	.071 (1.069)

Note. TC: Timed Contest Condition, DL: Difficulty Learning Condition.

As shown in Table 11, considering students with mild (level 1) ASD only (n = 6), the Wilcoxon signed-ranks test yielded a significant increase (p = .0261) in the number of math problems completed correctly per minute between baseline and intervention (M = 1.79 vs. M = 2.61), with a moderate effect size (ES = .51). Considering students with moderate (level 2) ASD only, the Wilcoxon signed rank test p value was only approaching significance (p = .063), and the very small sample size of this subgroup (n = 2) associated with increased Type II error prevents us from drawing definitive conclusions. There was a very large effect size (ES = 1.387) reflecting an increase in mean MLPR from baseline (M = 1.21) to intervention (M = 2.22). This increase was also significant for each of the strategy (see Table 12) with small to moderate effect sizes, among which the response modulation strategy yielded the highest effect size of 0.778 (p = .004) for improving the mathematical learning performance rates for students with ASD, followed by the attention deployment strategy (ES = 0.646, p = .004) and the cognitive change strategy (ES = 0.463, p = .012).

Table 11. Change in mathematical learning performance rates between baseline and intervention phases for each level of ASD

Level of ASD	Phases						*p*-value (ES)
	Baseline			Intervention			
	Median	Mean	Range	Median	Mean	Range	
Moderate	1.17	1.21	0.33; 2.17	2.44	2.22	0.88; 3.13	.063 (1.387)
Mild	1.40	1.79	0.17; 5.22	1.22	2.61	0.28; 6.93	.026* (0.510)

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

Table 12. Change in mathematical learning performance rates between baseline and intervention phases for the different strategies

Dependent variable	Phases						*p*-value (ES)
	Baseline			Intervention			
	Median	Mean	Range	Median	Mean	Range	
RM	1.32	1.65	0.17; 5.22	1.72	2.78	0.24; 8.50	.004* (0.778)
CC	0.85	1.64	0.17; 5.22	1.29	2.36	0.33; 7.71	.012* (0.463)
AD	2.09	2.60	0.33; 5.22	2.95	3.55	0.89; 6.00	.004* (0.646)

Note. $^{*}p < .05$. RM: Response Modulation, CC: Cognitive Change, AD: Attention Deployment.

6. Discussion and conclusion

This research proposed an e-learning model that features emotion recognition and regulation to enhance e-learning for students with ASD. An emotion recognition classifier, an emotion regulation method, and an emotion regulation model were then developed to realize the proposed e-learning model. To the best of our knowledge, this is the first model that integrates and performs domain-specific academic emotion recognition and emotion regulation. Based on this model, a mathematics e-learning environment was implemented, and experiments were conducted to verify the effectiveness of the proposed e-learning model and emotion regulation model. Next is a discussion of the research questions presented in Section 1:

6.1. Are emotion regulation strategies effective in reducing negative academic emotions of a student with ASD in the proposed e-Learning model?

Significant improvements in negative emotional behavior were observed between baseline and intervention phases (overall $p = .000^{***}$, TC $p = .004^{*}$, DL $p = .004^{*}$) with large effect size (overall ES = 1.010, TC ES = 0.888, DL ES = 1.155), both generally and separately in the two experiment conditions. These preliminary findings suggest that negative academic emotions of students with ASD can be improved through emotion regulation strategies embedded in the proposed e-Learning model.

6.2. Are emotion regulation strategies effective in improving the mathematical learning performance of the student with ASD in the proposed e-Learning model?

The overall improvement of mathematical learning performance was highly significant between baseline and intervention phases (overall $p = .005^{*}$) with moderate effect size (ES = .594), and the changes in the mathematical learning performance were also statistically significant between baseline and intervention phases in the *time contest* and *difficulty learning* conditions ($p = .004^{**}$) corresponding to large effect sizes (TC ES = .888, DL ES = 1.155). Despite the small sample size of this study, these preliminary findings suggest that the mathematical learning performance of students with ASD can be improved through emotion regulation strategies embedded in the proposed e-Learning model.

6.3. Are there differences in the effectiveness of emotion regulation strategies for students with mild and moderate ASD in the proposed e-Learning model?

For participants with mild (level 1) ASD, a significant difference was obtained in the percentages of negative emotional behaviors between baseline and intervention phases with a large effect size ($p = .000^{***}$, ES = .854). A significant difference was also obtained in the number of math problems completed correctly per minute between baseline and intervention phases with a moderate effect size ($p = .026$, ES = .51). That is, the emotion regulation strategies embedded in the proposed e-Learning model significantly decreased the percentages of negative emotional behaviors and also significantly increased the rates of mathematical learning performance among the students with mild (level 1) ASD. For participants with moderate (level 2) ASD, the mean percentage of negative emotional behaviors decreased from the baseline (65.67%) to intervention phases (25.26%) which was approaching significance ($p = .063$), whereas the mean rate of mathematical learning performance increased from the baseline (1.21) to intervention phases (2.22) which was also approaching significance ($p = .063$). This result could be linked to the very small sample size of this subgroup ($n = 2$), and replications in larger samples studies are needed.

It has generally been recognized that there are close interactions between mathematics learning and affective, motivational, and cognitive processes and their regulation (Hanin et al., 2017). Hannula (2019) had argued that the most common type of mathematics-related effect is emotions, which are considered to be important predictors of students' self-regulation and achievement (Pekrun, 2016). Emotions related to either achievement outcomes (e.g., anxiety) or learning-related activities (e.g., enjoyment of learning, anger at the task demands) have been termed as academic emotions (Frenzel, Pekrun, & Goetz, 2007; Pekrun, 2016). Previous research with typically developing students has shown that negative academic emotions, such as anger and anxiety, can impair a student's performance at complex or difficult tasks and his/her motivation to stay on mathematics assignments (Ma, 1999), and may reduce cognitive resources and self-regulation of learning (Ahmed, Minnaert, Van der Werf, & Kuyper, 2013). In contrast, positive academic emotions may have a positive influence on the use of flexible learning strategies (Pekrun, Goetz, Titz, & Perry, 2002), and can motivate students to work on mathematics problem-solving (Liljedahl, 2005). In a recent effort to enhance the understanding of the roles of academic emotions and emotion regulation in mathematics learning, Hanin and colleagues (Hanin et al., 2017) developed the Children's Emotion Regulation Scale in Mathematics (CERS-M). This indicates an increasing acknowledgment of the importance of domain-specific academic emotion regulation.

Students with ASD have significantly higher prevalence rates of emotional difficulties, including anger and anxiety, compared to typically developing peers (Totsika et al., 2011). The data from the study by Oswald and colleagues (Oswald et al., 2016) suggest that anxiety, one of the negative academic emotions, may serve as a potential target for intervention to enhance mathematics achievement of students with ASD. Although research exists documenting the relationship between these emotional difficulties (i.e., anxiety and anger) and deficits in emotion regulation in students with ASD (Fujii, Renno, McLeod, Lin, Decker, Zielinski, & Wood, 2013; Rieffe, Camodeca, Pouw, Lange, & Stockmann, 2012), there is still a need for more empirical investigation of domain-specific academic emotion regulation interventions. Taken together, albeit the small sample size and preliminary nature of the findings, the results of the current study indicate that the proposed e-learning model featuring emotion recognition and emotion regulation had moderate to large effects on reducing negative academic emotions and improving the mathematical learning performance of students with ASD. These preliminary findings support further evaluation of the efficacy of the intervention in a larger randomized control trial.

Due to the small sample size and data not normally distributed, a nonparametric Wilcoxon signed-rank test was used in the current study to examine the statistical significance of the difference between baseline and intervention conditions for all participants. That is, within-group comparison of dependent variables was done in this study using Wilcoxon signed rank test for comparing repeated measurements on a single sample to assess whether their population mean ranks differ. The results of this study demonstrated significant differences on almost all dependent variables. Nevertheless, it is important to note that a few participants with ASD exhibited not only high within-subject variability but also high between-subject variability. This can be in part explained by the substantial inter- and intraindividual variabilities in participants' state emotions, defined as the momentary context-specific appraisals, emotions, and strategies that emerge during a person-environment transaction (Schutz & Davis, 2000) and idiosyncratic emotions (Begeer, Koot, Rieffe, Terwogt, & Stegge, 2008), often observed in ASD. That is, one of the key challenges in developing domain-specific academic emotion recognition and emotion regulation interventions for students with ASD is that this group is highly heterogeneous (Wodka, Mathy, & Kalb, 2013), and abundant inter- and intraindividual heterogeneity is often observed in their progress over time in responses to interventions.

Our results highlight the importance of domain-specific adaptive academic emotion regulation interventions for students with ASD, and indicate the need for future studies aimed at further improving the adaptive intervention models with additional intervention strategies like emotion expression and situation selection (Hanin et al., 2017) or learning contexts, such as collaborative virtual environments, single-player and multiplayer educational games, and Virtual Reality (VR) and Augmented Reality (AR) environments.

Compared to other works on affective-related e-learning that focused merely on emotion recognition or emotion regulation (Bernard-Opitz et al., 2001; Sze & Wood, 2007; Lehmkuhl et al., 2008; Chorpita et al., 2009; Reaven, 2010; Cheng et al., 2010), this research integrated the methods of emotion recognition and emotion regulation to make the e-learning environment more effective and applicable to students with ASD. Besides helping students with ASD to learn effectively by reducing emotional interference on mathematical learning processes, the e-learning platform developed in this research can also help special education teachers and parents to better understand the emotional state of students with ASD and thus provide appropriate assistance.

Finally, this research could be improved in several ways. First, a sample that included more participants with moderate ASD would yield experimental results that are more representative. Second, since some studies have found increased rates of emotional symptoms for students with ASD after entering adolescence (Keith, Jamieson, & Bennetto, 2019), and most of the participants in this research were students with ASD in elementary schools, it is suggested that adolescents with ASD or students with other disabilities can be targeted in future studies. Third, other variables related to learning performance could be investigated in future studies including, for example, the degree of commitment to learning and the motivation to learn. Fourth, individuals communicate emotional information not only through facial expression but also body posture (Bijlstra, Holland, Dotsch, & Wigboldus, 2019). A mechanism for the recognition of emotional body postures might also be incorporated into the e-learning platform to increase recognition accuracy. Fifth, recent studies (Hu et al., 2019; Sung et al., 2019; Cakir et al., 2019) demonstrate the potential of using AR/VR technologies in e-Learning, using related methodologies to design materials of emotion regulation strategy with immersive experience might further improve the effectiveness of this study.

Acknowledgments

This work was supported by the grants by the National Science Council (100-2628-S-024-003-MY3) and the Ministry of Science and Technology, Taiwan, R.O.C. (104-2511-S-024-002) awarded to the first author. The authors are grateful to the anonymous reviewers for their constructive comments and suggestions on an earlier version of this manuscript, and would like to thank the children, their parents and teachers for making this study possible.

References

Afzal, S., & Robinson, P. (2011). Designing for automatic affect inference in learning environments. *Educational Technology & Society*, *14*(4), 21-34.

Ahmed, W., Minnaert, A., Van der Werf, G., & Kuyper, H. (2013). Emotions, self-regulated learning, and achievement in mathematics: A Growth curve analysis. *Journal of Educational Psychology*, *105*(1), 150-161. doi:10.1037/a0030160

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). doi:10.1176/appi.books.9780890425596

Ashburner, J., Ziviani, J., & Rodger, S. (2010). Surviving in the mainstream: Capacity of children with autism spectrum disorders to perform academically and regulate their emotions and behavior at school. *Research in Autism Spectrum Disorders*, *4*(1), 18-27. doi:10.1016/j.rasd.2009.07.002

Begeer, S., Koot, H. M., Rieffe, C., Terwogt, M. M., & Stegge, H. (2008). Emotional competence in children with autism: Diagnostic criteria and empirical evidence. *Developmental Review, 28*(3), 342-369. doi:10.1016/j.dr.2007.09.001

Bejerot, S. (2007). An Autistic dimension: A Proposed subtype of obsessive-compulsive disorder. *Autism, 11*(2), 101-110. doi:10.1177/1362361307075699

Bernard-Opitz, V., Sriram N., & Nakhoda-Sapuan S. (2001). Enhancing social problem solving in children with autism and normal children through computer-assisted instruction. *Journal of Autism and Developmental Disorders, 31*(4), 377-384. doi:10.1023/A:1010660502130

Bieberich, A. A., & Morgan, S. B. (2004). Self-regulation and affective expression during play in children with autism or Down syndrome: A Short-term longitudinal study. *Autism and Developmental Disorders*, *34*(4), 439-448. doi:10.1023/B:JADD.0000037420.16169.28

Bijlstra, G., Holland, R. W., Dotsch, R., & Wigboldus, D. H. J. (2019). Stereotypes and prejudice affect the recognition of emotional body postures. *Emotion, 19*(2), 189-199. doi:10.1037/emo0000438

Cakir, R., & Korkmaz, O. (2019). The effectiveness of augmented reality environments on individuals with special education needs. *Education and Information Technologies, 24*(2), 1631-1659. doi:10.1007/s10639-018-9848-6

Campos, J. J., Frankel, C. B., & Camras, L. (2004). On the nature of emotion regulation. *Child Development*, *75*(2), 377-394. doi:10.1111/j.1467-8624.2004.00681.x

Carey, E., Hill, F., Devine, A., & Szücs, D. (2015). The Chicken or the egg? The Direction of the relationship between mathematics anxiety and mathematics performance. *Frontiers in Psychology, 6*(33). doi:10.3389/fpsyg.2015.01987

Chen, C. M., & Lee, T. H. (2011). Emotion recognition and communication for reducing second-language speaking anxiety in a web-based one-to-one synchronous learning environment. *British Journal of Educational Technology*, *42*(3), 417-440. doi:10.1111/j.1467-8535.2009.01035.x

Cheng, Y., & Ye, J. (2010). Exploring the social competence of students with autism spectrum conditions in a collaborative virtual learning environment: The Pilot study. *Computers & Education*, *54*(4), 1068-1077. doi:10.1016/j.compedu.2009.10.011

Chorpita, B. F., & Daleiden, E. L. (2009). Mapping evidence-based treatments for children and adolescents: application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting and Clinical Psychology*, *77*(3), 566-579. doi:10.1037/a0014565

Colby, K. M. (1973). The Rationale for computer-based treatment of language difficulties in nonspeaking autistic children. *Journal of Autism and Childhood Schizophrenia*, *3*(3), 254-260. doi:10.1007/BF01538283

Conner, C. M., White, S. W., Beck, K. B., Golt, J., Smith, I. C., & Mazefsky, C. A. (2019). Improving emotion regulation ability in autism: The Emotional Awareness and Skills Enhancement (EASE) program. *Autism*, *23*(5), 1273-1287. doi:10.1177/1362361318810709

Conner, C. M., White, S. W., Scahill, L., & Mazefsky, C. A. (2020). The Role of emotion regulation and core autism symptoms in the experience of anxiety in autism. *Autism*, *24*(4), 931-940. doi:10.1177/1362361320904217

Cortes, C., & Vapnik, V. N. (1995). Support-vector networks, *Machine Learning, 20*(3), 273-297. doi:10.1023/A:1022627411411

Curilem, S. G., Barbosa, A. R., & De Azevedo, F. M. (2007). Intelligent tutoring systems: Formalization as automata and interface design using neural networks. *Computers and Education*, *49*(3), 545-561. doi:10.1016/j.compedu.2005.10.005

Czapinski, P., & Bryson, S. E. (2003). Reduced facial muscle movements in autism: Evidence for dysfunction in the neuromuscular pathway. *Brain and Cognition*, *51*(2), 177-179.

Dalkey, N., & Helmer, O. (1963). An Experimental application of the Delphi method to the use of experts. *Management Science*, *9*(3), 458-467. doi:10.1287/mnsc.9.3.458

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, *1*(2), 170-177. doi:10.1037/1082-989X.1.2.170

Farmer, M., & Oliver, A. (2005). Assessment of pragmatic difficulties and socio-emotional adjustment in practice. *International Journal of Language & Communication Disorders*, *40*(4), 403–429. doi:10.1080/13682820400027743

Floress, M. T., Zoder-Martell, K., & Schaub, R. (2017). Social skills plus relaxation training with a child with ASD in the schools. *Research in Developmental Disabilities*, *71*, 200-213. doi:10.1016/j.ridd.2017.10.012

Forsten, C. (2004). The Problem with word problem. *Principal*, *84*(2), 20-23.

Frenzel A. C., Pekrun, R., & Goetz, T. (2007). Perceived learning environment and students' emotional experiences: A Multilevel analysis of mathematics classrooms. *Learning and Instruction*, *17*(5), 478-493. doi:10.1016/j.learninstruc.2007.09.001

Fujii, C., Renno, P., McLeod, B. D., Lin, C. E., Decker, K., Zielinski, K., & Wood, J. J. (2013). Intensive cognitive behavioral therapy for anxiety disorders in school-aged children with autism: A Preliminary comparison with treatment-as-usual. *School Mental Health*, *5*(1), 25-37. doi:10.1007/s12310-012-9090-0

Georgiou, A., Soulis, S., Rapti, D., & Papanikolaou, K. (2018). Math anxiety of students with high functioning autism spectrum disorder. *American International Journal of Social Science, 7*(4). doi:10.30845/aijss.v7n4p15

Ginsburg, H. P. (1997). Mathematics learning disabilities: A View from developmental psychology. *Journal of Learning Disabilities, 30*(1), 20-33. doi:10.1177/002221949703000102

Golan, O., & Baron-Cohen, S. (2006). Systemizing empathy: Teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. *Development and Psychopathology*, *18*(2), 591–617. doi:10.1017/S0954579406060305

Grolnick, W. S., Bridges, L. J., & Connell, J. P. (1996). Emotional self-regulation in two-year-olds: Strategies and emotional expression in four contexts. *Child Development*, *67*(3), 928-941. doi:10.2307/1131871

Gross, J. J. (1998). The Emerging field of emotion regulation: An Integrative review. *Review of General Psychology*, *2*(3), 271-299. doi:10.1037/1089-2680.2.3.271

Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology*, *39*(3), 281-291. doi:10.1017/s0048577201393198

Gross, J. J., & Thompson, R. A. (2007). Emotion regulation: Conceptual foundations. In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 3-24). New York, NY: Guilford Press.

Guyon, I., & Elisseeff, A. (2003). An Introduction to variable and feature selection. *Journal of Machine Learning Research*, *3*(7-8), 1157-1182. doi:10.1162/153244303322753616

Hamilton, L. M. (2000). *Facing autism: Giving parents reasons for hope and guidance for help*. Colorado Spring, CO: Waterbrook Press.

Hanin, V., Grégoire, J., Mikolajczak, M., Fantini-Hauwel, C., & Van Nieuwenhoven, C. (2017). Children's Emotion Regulation Scale in Mathematics (CERS-M): Development and validation of a self-reported instrument. *Psychology*, *8*(13), 2240-2275. doi:10.4236/psych.2017.813143

Hannula, M. S. (2019). Young learners' mathematics-related affect: A Commentary on concepts, methods, and developmental trends. *Educational Studies in Mathematics, 100*(3), 309-316. doi:10.1007/s10649-018-9865-9

Heimann, M., Nelson K. E., Tjus, T., & Gillberg, C. (1995). Increasing reading and communication skills in children with autism through an interactive multimedia computer program. *Journal of Autism and Developmental Disorders*, *25*(5), 459-480. doi:10.1007/BF02178294

Hilton, C., Ratcliff, K., Collins, D., Flanagan, J., & Hong, I. (2019). Differences in flourishing for children with Autism Spectrum Disorder (ASD). *American Journal of Occupational Therapy*, *12*(6), 952-966. doi:10.5014/ajot.2019.73S1-RP301C

Hoover, J. J., & Patton, J. R. (1997). *Curriculum adaptation for students with learning and behavior problems: Principles and practices* (2nd ed.). Austin, TX: Pro-ed.

Hu, X., & Han, Z. R. (2019). Effects of gesture-based match-to-sample instruction via virtual reality technology for Chinese students with autism spectrum disorders. *International Journal of Developmental Disabilities*, *65*(5), 327-336. doi:10.1080/20473869.2019.1602350

Huang, C. L., Chen, M. C., & Wang, C. J. (2007). Credit scoring with a data mining approach based on support vector machines. *Expert Systems with Applications, 33*(4), 847-856.

Hughes, C., Russell, J., & Robbins, T. W. (1994). Evidence for executive dysfunction in autism. *Neuropsychologia*, *32*(4), 477-492. doi:10.1016/0028-3932(94)90092-2

Kapoor, A., Burleson, W., & Picard, R. W. (2007). Automatic prediction of frustration. *International Journal of Human-Computer Studies, 65*(8), 724-736. doi:10.1016/j.ijhcs.2007.02.003

Keith, J. M., Jamieson, J. P., & Bennetto, L. (2019). The Importance of adolescent self-report in autism spectrum disorder: Integration of questionnaire and autonomic measures. *Journal of Abnormal Child Psychology*, *47*(4), 741-754.

Khowaja, K., & Salim, S. S. (2013). A Systematic review of strategies and computer-based intervention (CBI) for reading comprehension of children with autism. *Research in Autism Disorders*, *7*(9), 1111-1121. doi:10.1016/j.rasd.2013.05.009

Kohavi, R. (1995). A Study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence 12* (pp. 1137-1143). San Francisco, CA: Morgan Kaufmann.

Konstantareas, M. M., & Stewart, K. (2006). Affect regulation and temperament in children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *36*(2), 143-154. doi:10.1007/s10803-005-0051-4

Kriete, T., & Noelle, D. C. (2015). Dopamine and the development of executive dysfunction in autism spectrum disorders. *PloS one*, *10*(3), e0121605. doi:10.1371/journal.pone.0121605

Lazarus, R. S., & Folkman, S. (1984). *Psychological stress and the coping process*. New York, NY: Springer.

Lehmkuhl, H. D., Storch, E. A., Bodfish, J. W., Geffken, G. R. (2008). Brief report: Exposure and response prevention for obsessive compulsive disorder in a 12-year-old with autism. *Journal of Autism and Developmental Disorders*, *38*(5), 977-981. doi:10.1007/s10803-007-0457-2

Lerner, J. W., & Kline, F. (2006). *Learning disabilities and related disorders-Characteristics and teaching strategies* (10th ed.). Boston, MA: Houghton Mifflin Company.

Leyfer, O. T., Folstein, S.E., Bacalman, S., Davis, N. O., Dinh, E., & Morgan, J. (2006). Comorbid psychiatric disorders in children with autism: Interview development and rates of disorder. *Journal of Autism and Developmental Disorders*, *36*(7), 849-861. doi:10.1007/s10803-006-0123-0

Liljedahl, P. G. (2005). Mathematical discovery and affect: The Effect of AHA! experiences on undergraduate mathematics students. *International Journal of Mathematical Education in Science & Technology*, *36*(2), 219-235. doi:10.1080/00207390412331316997

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education*, *30*(5), 520-540. doi:10.2307/749772

MacNeil, B. M., Lopes, V. A., & Minnes, P. M. (2009). Anxiety in children and adolescents with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *3*(1), 1-21. doi:10.1016/j.rasd.2008.06.001

Mao, X., & Li, Z. (2010). Agent based affective tutoring systems: A Pilot study. *Computers & Education*, *55*(1), 202-208. doi:10.1016/j.compedu.2010.01.005

Maughan, A. L., & Weiss, J. A. (2017). Parental outcomes following participation in cognitive behavior therapy for children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 47*(1), 3166-3179. doi:10.1007/s10803-017-3224-z

McLaughlin, V. L. (1993). Curriculum adaptation and development. In B. S. Billingsley (Ed.), *Program leadership for serving students with disabilities* (pp. 167-168). Richmond, VA: Virginia Department of Education.

Mehrabian, A. (1968). Communication without words. *Psychology Today*, *2*(9), 52-55.

Mendes, E., & Mosley, N. (2008). Bayesian network models for web effort prediction: A Comparative study. *IEEE Transactions on Software Engineering, 34*(6), 723-737. doi:10.1109/TSE.2008.64

Moore, M., & Calvert, S. (2000). Vocabulary acquisition for children with autism: Teacher or computer instruction. *Journal of Autism and Developmental Disorders, 30*(4), 359-362. doi:10.1023/A:1005535602064

Moridis, C. N., & Economides, A. A. (2008). Toward computer-aided affective learning systems: A Literature review. *Educational Computing Research*, *39*(4), 313-337. doi:10.2190/EC.39.4.a

Oswald, T. M., Beck, J. S., Iosif, A. M., McCauley, J. B., Gilhooly, L. J., Matter, J. C., & Solomon, M. (2016). Clinical and cognitive characteristics associated with mathematics problem solving in adolescents with autism spectrum disorder. *Autism Research*, *9*(4), 480-490. doi:10.1002/aur.1524

Parkinson, B., & Totterdell, P. (1999). Classifying affect-regulation strategies. *Cognition & Emotion*, *13*(3), 277-303. doi:10.1080/026999399379285

Patnam, V. S. P., George, F. T., George, K., & Verma, A. (2017). Deep learning based recognition of meltdown in autistic kids. In *Proceeding of the 2017 IEEE International Conference on Healthcare Informatics (ICHI)* (pp. 391-396). Park City, UT: IEEE. doi:10.1109/ICHI.2017.35

Pekrun, R. (2016). Academic emotions. In R. W. Kathryn, & B. M. David (Eds.), *Handbook of motivation at school* (pp. 120-144). Loudon, UK: Routledge Handbooks Online. doi:10.4324/9781315773384.ch7

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A Program of qualitative and quantitative research. *Educational Psychologist*, *37*(2), 91-105. doi:10.1207/S15326985EP3702_4

Picard, R. W. (1997). *Affective computing.* Cambridge, Massachusetts: MIT Press.

Pierce, K., & Courchesne, E. (2001). Evidence for a cerebellar role in reduced exploration and stereotyped behavior in autism. *Biological Psychiatry, 49*(8), 655-664. doi:10.1016/S0006-3223(00)01008-8

Politis, Y., Sung, C., Goodman, L., & Leahy, M. (2019). Conversation skills training for people with autism through virtual reality: using responsible research and innovation approach. *Advances in Autism, 6*(1), 3-16. doi:10.1108/AIA-05-2018-0017

Provost, F., & Fawcett, T. (2001). Robust classification for imprecise environments. *Machine Learning, 42*(3), 203-231. doi:10.1023/A:1007601015854

Quinlan, J. R. (1979). Discovering rules from large collections of examples: A Case study. In D. Michie (Ed.) *Expert Systems in the Microelectronic Age* (pp. 168-201). Edinburgh, Scotland: Edinburgh University Press.

Quoidbach, J., Mikolajczak, M., & Gross, J. J. (2015). Positive interventions: An Emotion regulation perspective. *Psychological Bulletin*, *141*(3), 655-693. doi:10.1037/a0038648

Rapp, J. T., & Gunby, K. (2016). Task interspersal for individuals with autism and other neurodevelopmental disorders. *Journal of Applied Behavior Analysis*, *49*(3), 730-734. doi:10.1002/jaba.319

Reaven, J. (2010). The Treatment of anxiety symptoms in youth with high-functioning autism spectrum disorders: Developmental considerations for parents. *Brain Research*, *1380*, 255-263. doi:10.1016/j.brainres.2010.09.075

Reaven, J., Blakeley-Smith, A., Leuthe, E., Moody, E., & Hepburn, S. (2012). Facing your fears in adolescence: Cognitive-behavioral therapy for high-functioning autism spectrum disorders and anxiety. *Autism Research and Treatment*, *2012*(2), 423905. doi:10.1155/2012/423905

Reaven, J., Blakeley-Smith, A., Nichols, S., Dasari, M., Flanigan, E., & Hepburn, S. (2009). Cognitive-behavioral group treatment for anxiety symptoms in children with high-functioning autism spectrum disorders: A Pilot study. *Focus on Autism and Other Developmental Disabilities, 24*(1), 27-37. doi:10.1177/1088357608327666

Rieffe, C., Camodeca, M., Pouw, L. B. C., Lange, A. M. C., & Stockmann, L. (2012). Don’t anger me! Bullying, victimization, and emotion dysregulation in young adolescents with ASD. *European Journal of Developmental Psychology*, *9*(3), 351-370. doi:10.1080/17405629.2012.680302

Rivard, M., & Forget, J. (2012). Verbal behavior in young children with autism spectrum disorders at the onset of an early behavioral intervention program. *The Psychological Record, 62*(2), 165-186. doi:10.1007/BF03395796

Roman, H. T. (2004). Why math is so important? *Teach Direction*, *63*(10), 16-18.

Roskam E. E. (1997). Models for speed and time-limit tests. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 187-208). New York, NY: Springer.

Rothbaum, F., Weisz, J. R., & Snyder, S. S. (1982). Changing the world and changing the self: A Two process model of perceived control. *Journal of Personality and Social Psychology, 42*(1), 5-37. doi:10.1037/0022-3514.42.1.5

Schiaffino, S., Garcia, P., & Amandi, A. (2008). ETeacher: Providing personalized assistance to e-learning students. *Computers and Education*, *51*(4), 1744-1754. doi:10.1016/j.compedu.2008.05.008

Schutz, P. A., & Davis, H. A. (2000). Emotions and self-regulation during test taking. *Educational Psychologist, 35*(4), 243-256. doi:10.1207/S15326985EP3504_03

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

Siegel, S. (1956). The Wilcoxon Matched-pairs Signed Rank Test. In S. Siegel, & N. John Castellan Jr. (Eds.), *Non-parametric statistics for the behavioral sciences* (pp. 75-83). New York, NY: McGraw-Hill.

Singh, N. N., Lancioni, G. E., Singh, A. D. A., Winton, A. S. W., Singh, A. N. A., & Singh, J. (2011). Adolescents with Asperger syndrome can use a mindfulness-based strategy to control their aggressive behavior. *Research in Autism Spectrum Disorders*, *5*(3), 1103-1109. doi:10.1016/j.rasd.2010.12.006

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science, 240*(4857), 1285-1293. doi:10.1126/science.3287615

Swezey, S. (2003). Autism and ICT: A Guide for teachers and parents. *Computers & Education*, *40*(1), 95-96. doi:10.1016/S0360-1315(02)00096-9

Sze, K. M., & Wood, J. J. (2007). Cognitive behavioral treatment of comorbid anxiety disorders and social difficulties in children with high-functioning autism: A Case report. *Journal of Contemporary Psychotherapy*, *37*(3), 133-143. doi:10.1007/s10879-007-9048-y

Tariq, U., Lin, K.-H., Li, Z., Zhou, X., Wang, Z., Le, V., Huang, T. S., Lv, X., & Han, T. X. (2012). Recognizing emotions from an ensemble of features. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 42*(4), 1017-1026. doi:10.1109/TSMCB.2012.2194701

Totsika, V., Hastings, R. P., Emerson, E., Lancaster, G. A. & Berridge, D. M. (2011). A Population-based investigation of behavioral and emotional problems and maternal mental health: Associations with autism spectrum disorder and intellectual disability. *Journal of Child Psychology and Psychiatry*, *52*(1), 91-99. doi:10.1111/j.1469-7610.2010.02295.x

Tsai, L. Y. (1999). Psychopharmacology in autism. *Psychosomatic Medicine*, *61*(5), 651-665. doi:10.1097/00006842-199909000-00008

Van-Bockstaele, B., Atticciati, L., & Hiekkaranta, A.P. (2020). Choose change: Situation modification, distraction, and reappraisal in mild versus intense negative situations. *Motivation and Emotion, 44*(4), 583-596. doi:10.1007/s11031-019-09811-8

Wodka, E. L., Mathy, P., & Kalb, L. (2013). Predictors of phrase and fluent speech in children with autism and severe language delay. *Pediatrics*, *131*(4), e1128-e1134. doi:10.1542/peds.2012-2221

Xenos, M. (2004). Prediction and assessment of student behaviour in open and distance education in computers using Bayesian networks. *Computers & Education*, *43*(4), 345-359. doi:10.1016/j.compedu.2003.09.005

Sun, D., Looi, C.-K., Yang, Y., & Sun, J. (2020). Design and Implementation of the Boundary Activity Based Learning (BABL) Principle in Science Inquiry: An Exploratory Study. *Educational Technology & Society*, *23* (4), 147–162.

Design and Implementation of the Boundary Activity Based Learning (BABL) Principle in Science Inquiry: An Exploratory Study

Daner Sun[1], Chee-Kit Looi[2], Yuqin Yang[3*] and Jin Sun[4]

[1]Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong SAR, China // [2]National Institute of Education, Nanyang Technological University, Singapore // [3]Faculty of Artificial Intelligence in Education, Hubei Research Center for Educational Informationization, Central China Normal University, China // [4]Department of Early Childhood, The Education University of Hong Kong, Hong Kong SAR, China // dsun@eduhk.hk // cheekit.looi@nie.edu.sg // yuqinyang0904@gmail.com // jinsun@eduhk.hk

[*]Corresponding author

(Submitted March 27, 2020; Revised May 17, 2020; Accepted November 10, 2020)

ABSTRACT: Guided by the Boundary Activity based Learning (BABL) principle, mobile technology-supported inquiry learning activities were implemented in a primary four science class in Hong Kong. An exploratory study was conducted to examine the effects of the BABL guided inquiry activities on students' learning performance and to explore how the key element, the boundary object, operated in different learning spaces. In the study, mixed research methods were used to evaluate students' conceptual understanding and their engagement in and attitudes toward BABL activities. The reciprocal interactions of students' cognition were qualitatively analyzed in terms of the forms and functions of boundary objects in the BABL environment. The results showed that students made significant improvements in conceptual understanding and were engaged in BABL activities. The study also revealed that the generation of abstract boundary objects, together with physical boundary objects, promoted students' learning and thinking as they shuttled between the classroom and the outside. This research contributes to informing educators about how to design and implement technology-supported teaching and learning through the use of boundary objects in crossing learning contexts.

Keywords: Boundary object, Boundary Activity based Learning (BABL) principle, Science inquiry, Crossing learning contexts

1. Introduction

Recently, a growing number of studies have investigated the relationship between learning in formal spaces and learning in informal spaces (Bonnette, Crowley, & Schunn, 2019). Many researchers have agreed that learning in informal spaces should not be viewed as an inferior form of learning but as a fundamental and valuable activity in its own right (Bell et al., 2009; Rogers, 2014). With the increasingly pervasive use of information and communications technology (ICT), wireless, mobile and ubiquitous technologies provide learners with various opportunities to link their experiences across multiple locations. Mobile learning has therefore been integrated into the teaching and learning of subjects (e.g., languages, math, science and music), from primary to university levels, both inside and outside of the classroom (Birch, 2017; Drigas & Pappas, 2015).

However, at a practical level, science learning in informal contexts receives less support than learning in formal contexts (Rogoff et al., 2016); consequently, collecting evidence of students' learning in informal contexts remains a challenge (Crompton, Burke, & Gregory, 2017). Much research in this area is descriptive and lacks a theoretical base (Anderson, Lucas, & Ginns, 2003). In the field of mobile learning, despite growing effort to create partnerships between schools and informal learning settings, the systematic documentation of such projects is limited (Shadiev et al., 2017). Moreover, few studies have been conducted concerning the structures required to support inquiry learning in informal contexts (Bai, 2019). As a result, there continue to be significant challenges in designing pedagogical learning scenarios in which learning takes place in both formal and informal spaces.

To address the abovementioned issues, many attempts have been done to improve learning synergy in formal and informal spaces through hybrid approaches (Lewin & Charania, 2018). Here we use learning in informal spaces which could accommodate the different patterns of learning activities, for example, informal learning (i.e., science museums, zoos and outdoor settings), out of classroom activities, and after-school programs (Tan, Jamaludin, & Hung, 2019). It fits the nature of the learning scenario in our study, where learning taking place in and out of the classroom but within the campus. Based on our year-long research efforts, we merge the merits of learning in formal and informal contexts via the notion of boundary objects, considering that few studies have explored the issue from the perspective of the generation of boundary objects. Boundary objects are generally defined as “entities that enhance the capacity of an idea, theory or practice to translate across culturally defined boundaries, for example, between communities of knowledge or practice” (Fox, 2011; Wenger, 1998) and have been applied in different fields (Fominykh et al., 2016). It is the first time we have used the notion of boundary objects to facilitate the connection of learning in different contexts through the design and generation of boundary objects in either abstract or physical patterns (Sun & Looi, 2018), in the field of mobile learning. In this study, we continued to consolidate the principle of boundary activity based learning (BABL) to guide the design and implementation of science inquiry learning supported by mobile technologies, with the objectives of promoting students’ learning, identifying the boundary objects that enable cohesive learning, both inside and outside of the classroom, and determining students’ cognitive transition trajectories in different learning spaces.

2. Theoretical framework

2.1. Facilitating the negotiation of science learning in formal and informal spaces

According to the hybrid view, formal learning and informal learning are not separate learning patterns: informal learning can take place in a formal learning context, and formal learning can take place in an informal context (Hofstein & Rosenfeld, 1996). Numerous studies have discussed the positive effects of connecting learning in formal and informal spaces and of creating structures to orchestrate learning activities in both contexts. As Falk and Balling (1982) noted, without orientation during outdoor learning activities, students are likely to focus on aspects of the environment that are irrelevant to learning. Gerber, Cavallo and Marek (2001) compared science learning in enriched and impoverished informal contexts. The study suggested that students improved their scientific reasoning skills the most when interacting with various informal learning contexts. Regarding the design of informal learning activities, researchers have offered valuable insights into the use of structures that can help to organize field trip activities better. For instance, Patrick, Mathews and Tunnicliffe (2013) suggested that field trips should incorporate problem-solving skills, be closely tied to the curriculum, focus on standards, and consider students’ needs. Sharples et al. (2014) proposed scripted learning methods of conducting outside-the-classroom inquiry activities.

Mobile learning offers new ways to extend education outside of the classroom and into the conversations and interactions of everyday life (Wang, Fang, & Miao, 2018). Mobile technologies possess a number of specific features that enable students to conduct authentic inquiry learning activities (Hwang & Tsai, 2011; Pu et al., 2016; Santos & Ali, 2012). Science projects created using mobile technology have demonstrated the efficacy of mobile learning as a means to improve students’ achievement, collaboration, motivation, and attitudes (Bano et al., 2018).

In brief, learning activities involving formal and informal contexts should be connected and well organized. Moreover, mobile technology is an effective tool for managing learning activities both inside and outside of the classroom. Regarding different learning settings, “sites of negotiation,” “boundaries,” and “border crossings” are useful and significant metaphors for understanding the mutual interaction of learning experiences and investigating the nature of learning (Aikenhead, 1996).

2.2. Foundations of the Boundary Activity-based Learning (BABL) principle

The BABL principle has been proposed and developed to guide the design of learning in multiple settings by responding to the negotiation and mutual interaction of cognition in crossing learning contexts. The BABL principle incorporates three components: (1) Boundary object: the boundary object is a prerequisite to designing boundary activities that ties together learning in and out of the classroom and monitors the learning process in informal spaces, in particular. The use of boundary objects allows learning activities in informal spaces to be integrated with learning activities from the standard curriculum. (2) Structure: the boundary activity is conducted in a pre-, during, and post-activity pattern to guarantee the continuum and stability of cognition or skills developed across the learning contexts. (3) Learning objectives: the learning objectives of the boundary activity should be defined based on the curriculum standard and the characteristics of the contextual variables in practice (see Sun & Looi, 2018; Sun & Looi, 2019). The underpinnings of the BABL principle are explained as follows.

2.2.1. Concept of the boundary object

Boundary objects are conceived as linkages that are plastic enough to adapt to local needs and the differing constraints of the various parties using them, yet robust enough to maintain a common identity across different sites. They can be any element that has the capacity to be understood by actors in more than one setting (Fox, 2011; Star & Griesemer, 1989). The creation and management of boundary objects are key processes in developing and maintaining coherence across intersecting social worlds (Star & Griesemer, 1989). Under this perspective, in examining issues of cross-cultural science education, Aikenhead (2001) proposed that learning about science is a cross-cultural event for most of students. Dealing with cognitive conflicts that arise from cultural clashes enables students to make sense of their learning, both inside and outside of the classroom.

Cognitive transition is considered smooth when its movement from one setting to another is harmonious and uncomplicated, with a merging of common sociocultural characteristics (Phelan, Davidson, & Cao, 1991). The smoother the cognitive transition between different learning contexts, the better the academic achievement and attitudes toward learning that students may attain (Costa, 1995). Based on this perspective, Akkerman and Bakker (2011) defined boundary objects as artifacts that articulate meaning, address multiple perspectives, and fulfil a bridging function. A boundary object can be either abstract or concrete (Cartwright & Mendell, 1984). It is a type of connecting link between communities of practice that can take the form of an artifact, document, term, concept, or other forms of ratification, around which communities of practice organize their interconnections (Wenger, 1998). Boundary objects in mobile learning scenarios are usually represented by concept maps, drawings, photos, videos, notes, or other relevant log data related to learning. However, they may take different forms as abstract patterns (Sun & Looi, 2018).

With respect to the BABL principle, boundary objects are the key elements in boundary activities. The investigation of students' cognitive transitions in different learning contexts mainly focuses on exploring the forms and functions of boundary objects in building students' knowledge and on exposing the mechanisms of boundary interaction.

2.2.2. The forms and functions of boundary objects

Various attempts have been made to explore the use of "boundary objects" in border-crossing contexts. Looi et al. (2009) investigated the occurrence of cognitive processes in a seamless learning context. One of the most significant processes in the operation of a cognitive system is the coordination between internal and external structures. Otero et al. (2011) proposed the use of external representations for understanding interconnections between individual cognitive processes, group processes, and the contributions that specific artifacts bring to the overall process. Wong, Chen, & Jan (2012) further emphasized the role of mediating artifacts in facilitating learners' effective transitions between scenarios in seamless learning. Several classes of mediating artifacts were identified: (1) subject matter artifacts, (2) physical artifacts, (3) socio-cognitive/non-physical artifacts, and (4) outcome artifacts. To foster community crossing interactions for sustained knowledge building, Zhang et al. (2018) proposed synthetic boundary objects in the form of idea thread syntheses, which indicated triggering students' deep thinking and reflection in different communities.

The above-mentioned studies share the view that establishing linkages between different learning contexts through boundary objects can improve cognitive interactions between individuals and collaborative groups. However, the functions of boundary objects vary according to the topic, subject, pedagogy, learning context, and technology. This observation of functional variability gives rise to further questions that motivated this study.

3. Research purposes and questions

We conducted an exploratory study with the following objectives: to investigate the impact of science inquiry guided by the BABL principle on students' learning performance, engagement, perceptions, and attitudes toward boundary activities inside and outside of the classroom, and to obtain insights into the forms and functions of boundary objects when students were engaged in reciprocal interactions of cognition in and out of the classroom. The following research questions were answered:

(1) To what extent did the students improve their conceptual understanding of science in the BABL science inquiry activities?

(2) How were the students engaged in reciprocal interactions of cognition in and out of the classroom?

(3) What were students' perceptions of and attitudes toward the boundary activities in the BABL science inquiry?

4. Methods

4.1. Participants and research contexts

As this investigation was the first pilot study to integrate the BABL principle into the Hong Kong Primary General Studies curriculum, small sample size was preferred (Johanson & Brooks, 2010). Consequently, a Grade 4 class of 37 students (10–11 years old) from a local primary school and their science teacher participated in this project, after giving their consent for data collection. The school's history of implementing e-learning and STEM education guaranteed that the participants could effectively adapt the proposed tools. Prior to the actual implementation of the BABL lessons, the teacher attended a professional development session facilitated by the first author. The power analysis showed that the sample size of 37 could achieve a power of .98, with an effect size of .95.

4.2. Lesson design for BABL guided science inquiry

In this project, a BABL lesson plan and corresponding teacher guide were developed on the topic of Energy, as prescribed by the Hong Kong Primary General Studies guidelines. As Table 1 shows, the BABL guided inquiry activities were scripted into four phases: (1) Context and Questions, (2) Investigation, (3) Sharing and Discussion, and (4) Conclusion (Bybee, 2009). To facilitate the smooth transition of cognition processes in and out of the classroom, structures were set in place for boundary activities in the form of pre-, during, and post-patterns according to a change in venue (Eshach, 2007). Besides, the learning objectives, as the integral components of the BABL principle, were articulated cohesively in the lesson plan to link learning in informal and formal spaces. Teacher and student interactions were mediated by boundary objects generated with the use of technological tools: nQuire-it (with Sense-it) and Schoology.

The nQuire-it learning system was developed by The Open University, UK. It enables teachers to create science projects and students to upload and comment on science data collected by mobile sensor toolkits using Sense-it. Figure 1 shows the basic interface of the Sense-it app and a generated real-time data chart. Schoology is a learning management system that enables teachers to upload teaching materials and design inquiry activities step by step, while monitor students' progress and learning processes in and out of the classroom. Figure 2 shows the design of inquiry learning activities supported by the BABL principle in Schoology. Android tablet mobile devices were provided to support Sense-it activities planned for outside of the classroom. In practice, two or three students shared one tablet.

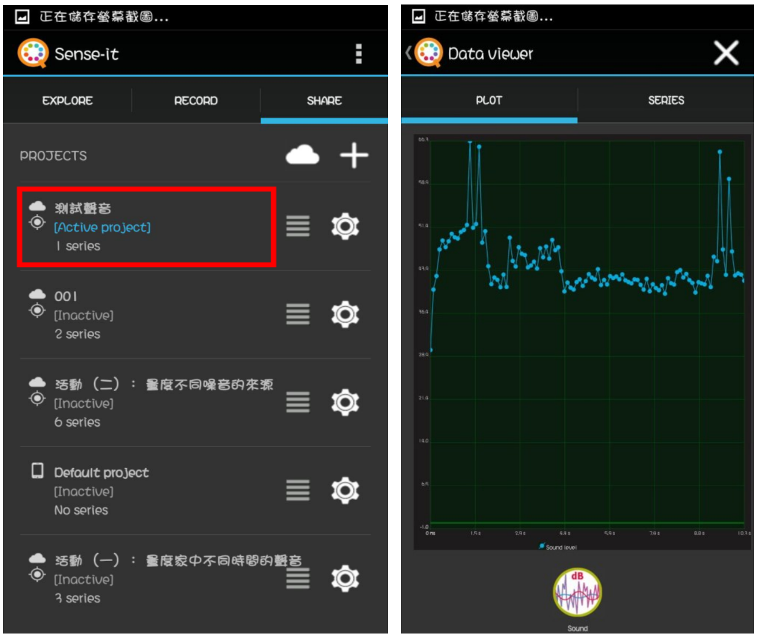


Figure 1. Sense-it app and data chart

Table 1. Lesson plan for BABL-guided inquiry

BABL inquiry topic: Energy – Solar power				
Inquiry phase	**Teacher activity**	**Student activity**	**Resources**	**Venue**
1. Context and Questions	• Introduction: solar energy and its application in daily life.	• Recognize solar energy.	Textbooks, websites, books and newspapers	Classroom (Pre-boundary activities)
2. Investigation	• Provide instruction on the inquiry task: exploring the intensity of sunlight. • Guide students to join the project in the nQuire-it platform. • Provide instructions on how to explore sunlight in the classroom. • Provide instructions on how to explore sunlight outside of the classroom.	• Work in groups: thinking about and discussing the task. • Join the project "Investigating sunlight" in nQuire-it. • Collect sunlight data in the classroom and upload the data to nQuire-it. • Discuss where, when and how to investigate sunlight outside of the classroom.	Sense-it: light sensor; Schoology Sense-it: light sensor	
	• Guide students to explore the strength of sunlight in groups outside of the classroom. • Preview/check students' work in nQuire-it and take note of some special cases, mistakes or deficiencies for real-time adjustment and support.	• Explore sunlight outside of the classroom. • Upload the sunlight data to nQuire-it.	Sense-it: light sensor; nQuire-it	Outside of the classroom: school garden, lobby, sports ground, balcony (During boundary activities)
3. Sharing & Discussion	• Guide students to review other groups' work and comment on it; highlight the work that receives the most positive/negative feedback. • Guide students to share reflections in Schoology.	• Review other groups' work. • Indicate "like" or "dislike" of others' work with comments. • Reflect on the experience and process of nQuire-it inquiry.	nQuire-it Schoology	Classroom (Post-boundary activities)
4. Conclusion	• Guide students to conclude the lesson and share their reflections.	• Draw conclusions and discuss the application of solar energy in daily life.	Schoology	

Figure 2. BABL-guided inquiry learning activities on Schoolog

4.3. Data collection and analysis

This study used one group pre-test–post-test quasi-experimental research design. Multiple types of data were collected to enable methodological triangulation (Fredricks et al., 2018). Figure 3 shows the data collection process.

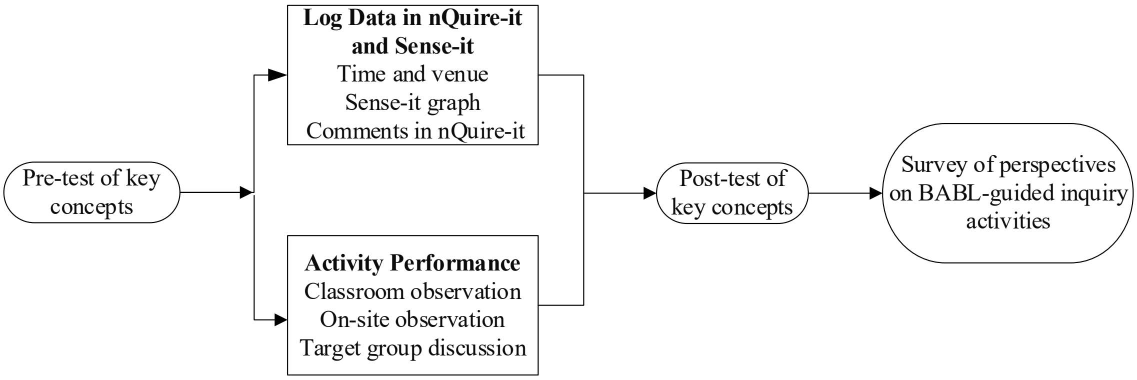


Figure 3. Data collection process

(1) Pre-and post-tests: To investigate students' conceptual understanding of key concepts, pre- and post-tests with identical test items were adopted. The test measured the understanding of the intensity of sunlight and the sources of sunlight in daily life. Altogether, five multiple-choice questions (two points each) and one open-ended question (five points) were included, for a total score of 15 points (20 mins). Q1 was concerned with the application of solar power in daily life. Q2, Q3, and Q5 were concerned with the times and locations of the strongest solar power. Q4 was concerned with ways to prevent overexposure to intense sunlight. Q6 was an open-ended question that required students to list the daily applications of solar power (two points) and to explain the mechanisms involved (three points). All of these items were drawn from standard academic tests and textbooks. An example question is

Q2. When is the sunlight the strongest?

A. Early Morning B. Noon C. Dusk D. Evening F. Other

The items' levels of difficulty and cognitive domains were based on the six levels of the revised Bloom's Taxonomy (Anderson et al., 2001). They were identified as Level 2 – Comprehending/Understanding: item 1; Level 3 – Applying: items 2 and 5; Level 4 – Analyzing: items 3 and 4; and Level 6 – Evaluating: item 6. To ensure the reliability and validity of the test items, two researchers and two experienced teachers reviewed and discussed whether the questions were appropriate. Each teacher had more than six years of science teaching experience. The Cronbach alpha's score was .845, indicating good internal consistency of the test. For the analysis of the pre-and post-test scores, a descriptive statistic was applied to summarize the overall results and to assess the normality of the tests. A paired sample *t*-test and an item analysis were used to compare the differences between the pre- and post-test results and between the students' responses to the questions.

(2) Students' log data and comments: To uncover the students' reciprocal interactions of cognition in and out of the classroom, log data (including time and date, location, data charts and comments) in nQuire-it was analyzed and served as the key body of evidence for profiling the boundary interactions experienced by students in the two learning contexts (Patten, Arnedillo Sánchez, & Tangney, 2006). Figure 4 shows the location and date of data collection presented in the nQuire-it system.

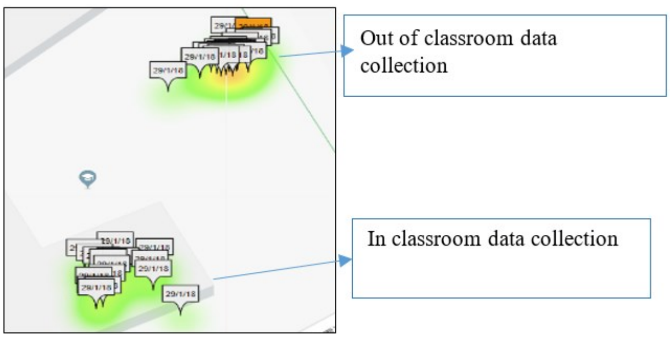
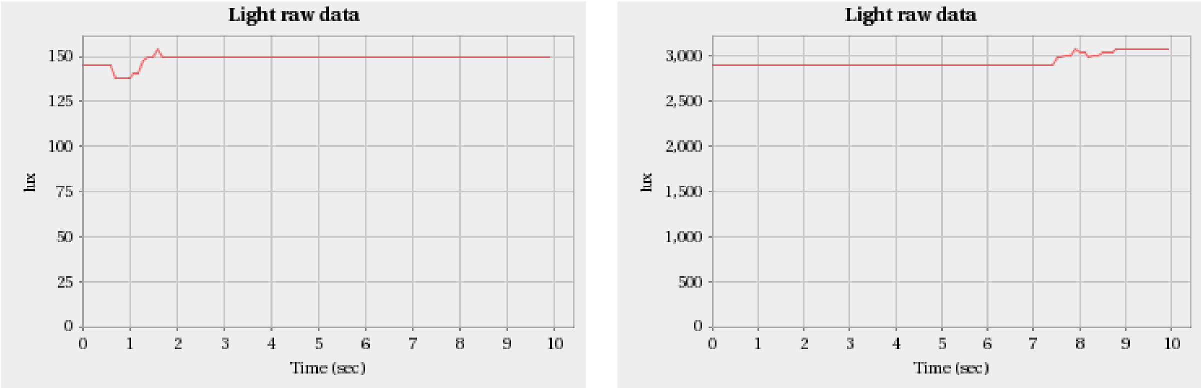


Figure 4. Students' data collection inside and outside of the classroom, using Sense-it

Moreover, as an important artifact of the physical boundary objects, the data charts in and out of the classroom were calculated in terms of the totals and means generated by each group. For example, two data charts on raw light, collected by the light sensors in the classroom and in the school garden, are shown in Figures 5 (a) and (b).



(a) (b)

Figure 5. (a) Light data collected in the classroom, (b) Light data collected in the school garden

The students' comments on their classmates' data charts were retrieved in the system and reviewed qualitatively as valuable evidence for assessing their reciprocal interactions of cognition in terms of thinking and understanding where abstract boundary objects in the representation of new concepts or understanding might be generated (Stahl, 2000). The coding and analysis were conducted independently by the first author and one researcher, and the inter-rater reliability coefficients were calculated as .85 for the data charts analysis and .88 for the comments analysis. Both coefficients indicated good inter-rater reliability.

(3) Observations and discourses: Three groups with two students for each were selected as targets from 17 groups. To ensure the same conditions of the groups (i.e., gender, collaboration skills, performance in normal classes), group characteristics were further discussed and identified by the teachers. Observation of target groups is a principal means of collecting qualitative data, as it enables researchers to observe a certain amount of interaction and achieve in-depth understanding of a topic in a limited time (Morgan, 1997). The researchers focused their observations of the three target groups on the groups' levels of engagement and reciprocal interactions of cognition during boundary activities.

Specifically, students' group discussions were videotaped, annotated, and transcribed in qualitative terms (Derry et al., 2010). Discussion is a common language for negotiating meaning across the boundaries of two learning contexts (Fominykh, et al., 2016). Two dimensions of coding methods were used (Lee & Irving, 2018) that involved discussions of boundary objects in three stages of boundary activities: (1) boundary object forms, e.g., abstract boundary objects and physical boundary objects, and (2) boundary object functions, in terms of cognitive transition. The inter-rater coding consistency of the qualitative data reached 93%, on average, which indicated good reliability.

(4) Survey of students' perceptions of and attitudes toward BABL guided inquiry activities. A survey was administered to investigate the students' perceptions of BABL guided inquiry activities, especially regarding the connections between formal learning and learning in informal spaces. The survey consisted of 12 questions on a 5-point Likert scale. The survey's dimensions were based on the three dimensions used in the Science Outdoor Learning Environment Inventory (Orion et al., 1997) that focus on (1) Environmental interaction: Interaction in outside activities (Q1, Q7); (2) Integration: The connection between in-class and out-of-class learning (Q2, Q3, Q4, Q5, Q6); and (3) Preparation and organization: Organization (Q8); and one dimension of Learning involvement (Brom et al., 2017): Motivation for outside activities (Q9, Q10, Q11, Q12). An example item of the 5-point Likert scale is *Q1. I discuss our tasks with my group members during outdoor inquiry activities.* For each dimension, the researchers designed the question items to fit the nature of the study and the topic. After discussions with two researchers and one experienced teacher, the survey was moderately revised, and the Cronbach alpha score was .860, indicating good internal consistency (Taber, 2018). A descriptive data analysis of the survey results was conducted to expose differences in the proportion of students' agreement with individual items.

5. Results

5.1. Changes in students' conceptual understanding

We first checked the normality distribution of the test scores in the pre- and post-tests using a q-q plot. The results showed that both test scores followed the normal distribution. The descriptive statistics of the pre- and post-test scores revealed a general improvement in the students' conceptual understanding after participating in the BABL guided inquiry activities. The results of a paired sample t-test indicated that such improvement was not only observable but significant (Mpre = 9.89, SDpre = 3.30; Mpost = 10.88, SDpost = 2.36; $t(36) = -2.939$, $p = 0.01$) (Table 2).

Table 2. Results of the paired sample t-test

Test	n	M	SD	t	p
Pre-test	37	9.89	3.30	-2.939	.006*
Post-test	37	10.88	2.36		

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

The results also showed a substantial increase in the percentage of correct answers. There was a 15% increase in correct answers to Q1 (Level 2) and 20% and 24% increases in correct answers to Q2 and Q5 (Level 3), respectively. For Q3 and Q4 (Level 4), the numbers of correct answers rose by 31% and 15%, respectively. In the open-ended question, the percentage of correct responses rose from 56% in the pre-test to 85% in the post-test. These results suggest that most of the students achieved a higher level of cognition concerning the key concepts involved in this topic.

5.2. Mechanisms of students' reciprocal interactions of cognition in and out of the classroom

5.2.1. Student engagement: Analysis of log data in nQuire-it and Sense-it and target groups' performance

The analysis of the data charts generated in the nQuire-it system suggested that nearly 90% of the students used the light sensor to collect data, both inside and outside of the classroom. This finding suggested that most students were engaged in the BABL activities. In the Investigation phase, all 17 groups collected light intensity data and uploaded the data charts to nQuire-it. A total of 89% of students successfully collected data in a collaborative manner, and 62.5% generated data charts for both inside and outside of the classroom. A descriptive data analysis showed that the average number of data charts generated by each group was 2.94, with the totals for the groups varying between one and nine (47 in total). In addition, the numbers of in-classroom data charts (i.e., 23) and out-of-classroom data charts (i.e., 24) were almost equal, indicating that the students performed equally well both in-classroom and out-of-classroom, without preference for the activity venue.

When the students returned to the classroom during the Sharing and Discussion phase, they reviewed their classmates' data charts through nQuire-it and indicated whether they "liked" or "disliked" them. On average, each chart received four "likes" and two "dislikes." The fact that each chart was reviewed by six students, on average, implied the students' active participation in peer review and assessment. More importantly, most students not only gave a general, summative assessment (such as a "like" or "dislike") but also provided formative feedback, which either justified their comments or represented their own interpretations of the data charts. Altogether, 30 students commented on their classmates' work, providing 25 positive comments and five negative ones. One student was especially active in commenting on his classmates' work. He noted that a sharply curved data chart meant "it [sunlight] changes a lot" and that the charts showed him "it's too dark" in the classroom, but "it [sunlight] is very bright" in the school garden. Such observations showed that peer comments on physical boundary objects could engage students in conceptual understanding and reasoning. Importantly, such comparison of the charts in and out of classroom might trigger the generation of abstract boundary objects.

During classroom activities, despite performing differently, most students were highly engaged. For some students, this activity was stimulating and motivating, as it was different from traditional science activities (Figure 6). As the activity unfolded, the three target groups were found to be interacting in effective ways. In Group 1, one student asked when the group could go outside for data collection, as she considered the classroom was not bright enough. This question implied that the student was motivated and engaged in her own process of reasoning. She wished to investigate more about the data outside. This self-reflection was probably the venue of generating abstract boundary objects (Pimmer, 2016). Group 2 attempted to block the light using textbooks during their data collection, as they wanted "to see whether the value (of light intensity) would be smaller." They tested it and were quite happy that their "guess was correct." In Group 3, the students were excited to discuss and share their data chart with the other groups (as illustrated in Figure 7). We transcribed and excerpted their discussion as follows:

Sa: What does your chart look like? Let me have a look.
Sb: Wait a moment; we need a few seconds. Okay, here you are.
Sa: The light looks very stable here, but mine is unstable. Mine is a little curved. Yours is flat.
Sc: I think you may be moving the tablet when you are using the app.
Sd: Yes, I agree with you. Please don't move next time.
Sa: I will do the test again. Let's see whose chart is more stable.

The above discourse emerged spontaneously, rather than being encouraged by the teacher or researchers. When students were involved in group comparison and reasoning, their motivation and engagement were triggered naturally.

Figure 6. Students engaged in the nQuire-it activity

Figure 7. Two groups discussing their findings

5.2.2. Reciprocal interactions of students' cognition in BABL guided inquiry activities

Further efforts were made to describe students' learning trajectories throughout the holistic learning experience, with a focus on capturing the forms and functions of boundary activities, especially for invisible and unexpected boundary objects in the form of abstract boundary objects during reciprocal interactions of students' cognition in crossing borders. According to data analysis of the three target groups' discussions, although the students were given the same instructions at the pre-, during, and post-boundary stages, they reacted in quite different ways. Table 3 provides a summary of students' foci of discussion related to the forms and functions of boundary objects at different stages of boundary activities. The first column indicates the forms and functions of boundary objects identified during the group discussions. The checkmarks represent the presence of boundary objectives.

It was found that most discussions concentrated on the functions of boundary objects that linked the application and transformation of the related knowledge, especially for during and post-boundary activities. These discussions were likely to generate abstract boundary objects we did not anticipate in the lesson design. Students who discussed the use of nQuire-it and Sense-it in their pre-boundary activity went directly to play with the light sensors in the Sense-it activity, as they were already familiar with the tools. They were interested in generating and observing the line charts (physical boundary objects) that described the intensity of light, and they became very curious (triggered by abstract boundary objects) when the shapes of the charts changed as they moved around. While investigating the light intensity in the classroom, they waved their pads or used their textbooks to block the light (control variables). They also carefully examined whether the actual charts fit their expectations (hypothesis-verification). With the classroom data in hand, some students were eager to collect data outside the classroom for comparison (comparison).

Table 3. Students' discussions inside and outside of the classroom

Discussion	Pre-boundary activities	During boundary activities	Post-boundary activities
Use of tools (the medium of boundary objects)	√	-	-
Observation of data charts (function of boundary objectives)	√	√	√
Shape of charts (function of boundary objects)	√	√	√
Comparison of charts (function of boundary objects)	-	√	√
Expectation of more comparisons (function of boundary objects)	-	√	√

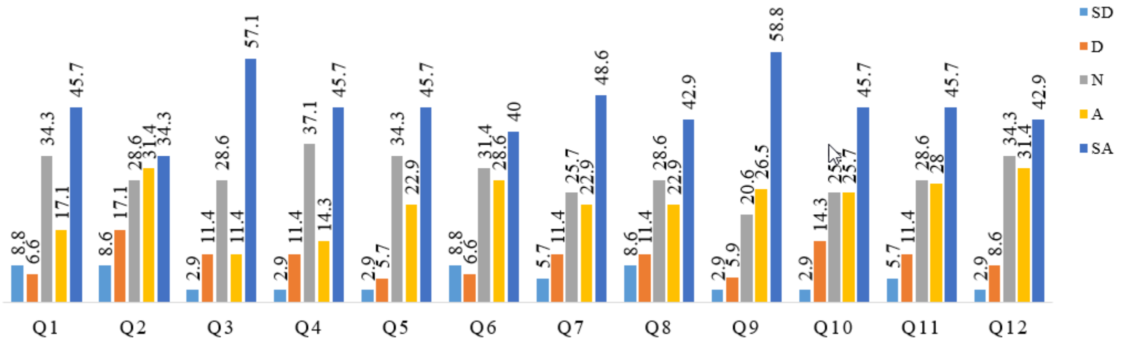
During the Sharing and Discussion process, they became deeply thoughtful as they compared and commented on each other's data charts (review and comment). These observed data exposed the forms of abstract boundary objects (i.e., control variables, hypothesis-verification, comparison, review, and comment) and the functions of boundary objects in students' knowledge transformation, engagement, and reasoning in crossing learning contexts, which were not anticipated in the lesson plan.

Specifically, some interesting and encouraging observations were captured, as follows. (1) Target Group 1 was determined to generate a "better" chart, in which the value of the intensity of sunlight would be higher and represented by a stable curve. To achieve this, they moved to the other side of the school garden, where the sunlight was stronger. (2) Target Group 2 planned to conduct several tests, compared the data charts generated in different locations, and took turns to collect data. (3) When the students were asked, "How do you connect your activities inside and outside of the classroom?" two of the six students mentioned they wanted to compare the intensity of sunlight collected outside with that collected in the classroom. One student said, "I know they are different because the source of light is different. One is light, and the other is sunlight." Another student further explained, "I can see that the value of the chart in the classroom is less than that of the one outside the classroom, because sunlight is a very powerful form of energy." Therefore, students elaborated their conceptual understanding and deepened their thinking when interacting with boundary objects (i.e., charts, control variables, comparison, concepts of sources of light) in and out of the classroom.

5.3. Students' perceptions of and attitudes towards boundary activities in BABL guided science inquiry

The proportion of student agreement on perceptions of and attitudes toward boundary activities was analyzed, as Figure 8 shows. The sum of the percentages of "agree" and "strongly agree" was calculated to represent students' main selection of each item. The first batch of survey questions concerned the relationship between learning inside and outside of the classroom (i.e., Q2, Q3, Q4, Q5, and Q6). As indicated by the students' responses to Q3 (68.5%), more than half of the students agreed that they had more opportunities to elaborate conceptual understanding developed in the classroom and relate it to their outdoor activities. Q4 and Q5 concerned whether knowledge attained outside could help students' learning in the classroom and whether the out-of-classroom activities could improve their thinking. Once again, the responses were positive, with 60% for Q4 and 68.6% for Q5. The proportion of positive responses to Q2 was also above 60% (65.7%), meaning that those students agreed that the out-of-classroom inquiry was related to their learning in the classroom and the data collected outside could be studied in the classroom later.

The second batch of survey questions examined the students' attitudes toward the post-boundary activity (i.e., Q1, Q7, and Q8). On average, over 60% of the students reflected that the teacher reviewed the data they collected in the outdoor activities and then discussed it in the classroom, which provided them with an opportunity to learn from their classmates. The final batch of survey questions (i.e., Q9, Q10, Q11, and Q12) explored the students' attitudes toward BABL guided inquiry. It shows that 85.3% of the students agreed the out-of-classroom activities done with their partners (Q9), 71.4% thought they had made a great effort in the inquiry activities (Q10), 73.7% felt they were engaged in the mobile learning activities (Q11) and 74.3% said they would like to use Sense-it to do outdoor activities when studying other science topics (Q12). These results suggest that students had a positive attitude toward the BABL inquiry activities.



*SD-Strongly disagree, D-Disagree, N-Neutral, A-Agree, SA-Strongly agree

Figure 8. Students' responses to the survey questions

6. Discussion and conclusion

Using mixed research methods, this exploratory study presents the impacts of BABL principle guided science inquiry activities on students learning performance and engagement and clarifies the mechanisms of students' reciprocal interactions of cognition in crossing learning contexts. In this study, learning takes place inside and outside of the classroom, on campus, to create scenarios of learning in formal and informal spaces.

6.1. Impact on students' conceptual understanding

To answer RQ1, the statistical analysis of the pre- and post-test results confirmed the students' changes of conceptual gains before and after the intervention. The analysis of the question items showed students' general improvement and achievement at different levels of cognition. Moreover, students' responses to the open-ended question became more active and accurate. These results offer strong evidence of the effectiveness of BABL learning in students' science learning.

6.2. Mechanisms of the students' reciprocal interactions of cognition in and out of the classroom

Log and discourse data were collected and analyzed to answer RQ2. Based on the analysis of the log data, students' engagement was verified by their active participation in nQuire-it activities during the Investigation and Sharing and Discussion phases. Not only did the students generate data charts with great zeal both inside and outside of the classroom, they also showed a great willingness to review and comment on their classmates' work, as was reflected in their discourse during the data collection. These results echo the finding that peer feedback is intended to actively and deeply engage students in learning (Moore & Teather, 2013). Meanwhile, the shared artifacts and methods contributed to providing the capacity to negotiate interests and transform knowledge, as reported in relevant studies (Carlile, 2004; Polman & Hope, 2014).

To clarify the mechanism of reciprocal interactions of cognition based on boundary objects, some valuable and unexpected findings were obtained. Prior to implementing the BABL learning activities in a real context, we expected the generation of boundary objects, in the form of either physical or abstract boundary objects, would improve the interaction and negotiation of students' cognition between formal and informal contexts. In the lesson plan, these boundary objects consisted primarily of physical artifacts (i.e., data charts, comments) generated inside and outside of the classroom, and we did not know what the abstract boundary objects would be. However, the study found that abstract boundary objects were also created and leveraged by the students in the form of conceptual understanding, scientific methods (i.e., hypothesis-verification, comparison, control variables) and thinking skills (i.e., reflection). The generation of abstract boundary objects accompanied the physical boundary objects. Our findings indicates that the boundary objects have different meanings and identities in different social worlds that they inhabit, and they are dynamic and have emergent characteristics (Pimmer, 2016).

Students' comments regarding the data charts expressed their further thinking about the reasons for data differences across different learning contexts. As mentioned above, one student provided explanations and conducted a critical reflection on why three different data charts were generated inside and outside of the classroom. This observation shows that the students' critical reflection as an abstract boundary object promotes knowledge construction inside and outside of the classroom context (Fominykh et al., 2013; Wang, Woo, & Zhao, 2009).

In analyzing the students' discussions, it was found that most focused on the functions of boundary objectives. We determined that the students formed hypotheses regarding the shapes of data charts, based on comparisons between the conditions of data collection inside and outside of the classroom. Such hypotheses served as abstract boundary objects for students to elaborate their understanding of the sources of light and the factors of sunlight intensity, which we did not expect prior to the BABL activities. These types of findings suggest that knowledge is not static and that the tension involved in the interaction between mediational boundary objects (whether they are physical or abstract) and the individuals using them tends to result in a continuous process of knowledge transformation and creativity (Wertsch & Rupert, 1993).

Moreover, our findings built on previous discussions of the roles of boundary objects in the knowledge transition process and how different types of cognitive transition can be triggered by diversified forms of boundary objects (Marheineke, Habicht, & Möslein, 2016). Such findings suggest that reciprocal interactions of cognition between learning inside and outside of the classroom do not rely solely on physical boundary objects. Rather, abstract boundary objects, generated by cross-referencing formal learning and learning in informal spaces, can mediate such reciprocal interactions.

6.3. Students' perceptions of and attitudes toward boundary activities in BABL guided science inquiry

The survey results indicate that most of the students believed that the BABL environment made them more collaborative and enabled them to connect the knowledge and skills they gained inside and outside of the classroom. They expressed a desire to continue using nQuire-it and Sense-it when studying other science topics. Similar results that the mobile learning experience and its external representations (ERs) can help to engage students in deep learning have been shown (Choi, Land, & Zimmerman, 2018). The more that people's innate psychological need for relatedness between different learning contexts is fulfilled, the more their motivation to learn will be triggered (Zainuddin, 2018).

To conclude, the consideration of the BABL principle in science inquiry supported by mobile technology in this study is within the domain of "science as culture" theory (Aikenhead et al., 1996), which postulates that boundary interaction is, in fact, the interaction between communities related to the learners' science learning. In seeking to identify the key episodes for realizing smooth transitions between different contexts, and in borrowing insights from relevant studies of boundary crossing (Hung & Chen, 2007), while incorporating the insights provided by Johnson's (1995) exploration of the stages and transitions in cognitive development, we suggest metaphorical terms for two patterns of cognitive transition in BABL environment. These patterns are "horizontal cognitive transition" and "vertical cognition transition." The term "horizontal cognitive transition" refers to a continuum of knowledge in different learning contexts. The term "vertical cognitive transition" is more related to knowledge elaboration or progression, in which deep, critical thinking or new understanding in a formal learning context is triggered by boundary objects generated in informal learning spaces. This type of transition can involve the formation of ideas, thoughts, hypotheses, or understanding in the informal learning context. It can also be explained by Bench's (1999) view that "[m]ost current accounts of learning transfer attribute cause or agency for the process to the abstraction and representation of knowledge by individual minds (p. 108)." We expect that the frequency of these two types of cognitive transition is closely related to students' learning achievements and cognition development in border-crossing contexts (Tsurusaki et al., 2012). Although not all of the groups in our study followed this learning trajectory, cognitive transition was salient in the BABL guided inquiry activities, and this type of interactive learning deserves further investigation.

7. Limitations, implications, and further research

This study has some limitations. First, it was an exploratory study and involved only a small group of learners. It did not apply an experimental design with a control-experimental group comparison. However, as an exploratory study, this investigation had its own merits, as it demonstrated the potential of the BABL principle for guiding the design and implementation of inquiry learning activities. Since many studies have considered models of learning designs with a specific focus on facilitating interactions between out-of-classroom learning and in-classroom learning, our lesson design and research findings contribute to the field of mobile learning in border-crossing learning contexts. These findings may inspire educational researchers and practitioners to reinforce the connections between in-class and out-of-class learning through the use of boundary objects and to improve cognitive interactions between learning in both formal and informal spaces.

Future investigations with a rigorous experimental design will be conducted to study the ways that learning switches between in-classroom and out-of-classroom settings and how students can achieve cognitive transformation through the generation of abstract boundary objects, or other forms of boundary objects.

Acknowledgement

The research is part of project work "Designing Science/STEM Learning with Boundary Activity based Learning (BABL) Principle" funded by Internal Research Grant, The Education University of Hong Kong. We would like to thank Professor Mike Sharples from Educational Technology in the Institute of Educational Technology at The Open University, UK, for kindly supporting our nQuire-it implementation and research in Hong Kong.

References

Anderson, D., Lucas, K. B., & Ginns, I. S. (2003). Theoretical perspectives on learning in an informal setting. *Journal of Research in Science Teaching, 40*(2), 1036-1054.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). *A Taxonomy for learning, teaching, and assessing: A Revision of Bloom's taxonomy of educational objectives* (Complete edition). New York, NY: Longman

Akkerman, S. F., & Bakker, A. (2011). Boundary crossing and boundary objects. *Review of Educational Research*, *81*(2), 132–169.

Aikenhead, G. S. (2001). Students' ease in crossing cultural borders into school science. *Science Education*, *85*(2), 180-188.

Aikenhead, G. S. (1996). Science education: Border crossing into the subculture of science. *Studies in Science Education, 27*(1), 1-52.

Bai, H. (2019). Pedagogical practices of mobile learning in K-12 and higher education settings. *TechTrends*, *63*, 611–620

Bano, M., Zowghi, D., Kearney, M., Schuck, S., & Aubusson, P. (2018). Mobile learning for science and mathematics school education: A Systematic review of empirical evidence. *Computers & Education*, *121*, 30-58.

Bell, P., Lewenstein, B., Shouse, A. W., & Feder, M. A. (2009). *Learning science in informal environments: People, places, and pursuits.* Washington, DC: The National Academy of Sciences.

Bench, K. (1999). Consequential transitions: A Sociocultural expedition beyond transfer in education. *Review of Research in Education*, *42*(1), 101–139.

Birch, H. J. S. (2017). Potential of SoundCloud for mobile learning in music education: A Pilot study. *International Journal of Mobile Learning and Organization*, *11*(1), 30-40.

Bonnette, R. N., Crowley, K., & Schunn, C. D. (2019). Falling in love and staying in love with science: Ongoing informal science experiences support fascination for all children. *International Journal of Science Education*, *41*(12), 1626-1643.

Brom, C., Děchtěrenko, F., Frollová, N., Stárková, T., Bromová, E., & D'Mello, S. K. (2017). Enjoyment or involvement? affective-motivational mediation during learning from a complex computerized simulation. *Computers & Education*, *114*, 236-254.

Bybee, R. W. (2009). The BSCS 5E instructional model and 21st century skills. *The National Academies Board on Science Education*. Retrieved from https://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_073327.pdf

Cartwright, N., & Mendell, H. (1984). What makes physics' objects abstract? In J. Cushing, C. Delaney & G. Gutting (Eds.), *Science and Reality* (pp. 134-152). Notre Dame, IN: University of Notre Dame Press.

Carlile, P. R. (2004). Transferring, translating, and transforming: An Integrative framework for managing knowledge across boundaries. *Organization Science*, *15*(5), 555-568.

Choi, G. W., Land, S.M., & Zimmerman, H. T. (2018). Investigating children's deep learning of the tree life cycle using mobile technologies. *Computers in Human Behavior*, *87,* 470-479.

Crompton, H., Burk, D., & Gregory, K. H. (2017). The use of mobile learning in PK-12 education: A Systematic review. *Computers & Education, 110*, 51-63.

Costa, V. B. (1995). When science is “another world”: Relationships between worlds of family, friends, school, and science. *Science Education*, *79*(3), 313-333.

Derry, S., Pea, R.D., Barron, B., Engle, R. A., Erickson, F., Goldman, R., Hall, R., Koschmann, T., Lemke, J. L., Sherin, M. G., & Sherin, B. L. (2010). Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. *Journal of Learning Sciences*, *19*(1), 3-53.

Drigas, A. S., & Pappas, M. A. (2015). A Review of mobile learning applications for mathematics. *International Journal of Interactive Mobile Technologies*, *9*(3), 18-23.

Eshach, H. (2007). Bridging in-school and out-of-school learning: Formal, non-formal, and informal education. *Journal of Science Education and Technology*, *16*(2), 171-190.

Falk, J. H., & Balling, J. D. (1982). The Field trip milieu: Learning and behaviour as a function of contextual events. *Journal of Educational Research*, *76*(1), 22-28.

Fominykh., M., Prasolova-Førland, E., Divitini, M., & Petersen, S. A. (2016). Boundary objects in collaborative work and learning. *Information System Frontiers*, *18*, 85-102.

Fominykh, M., Prasolova-Forland, E., Divitini, M., & Petersen, S. A. (2013). Hands-on learning of cooperation technology: Combining knowledge construction and reflection. In M. J. W. Lee (Ed.), *Proceedings of International Conference on Teaching, Assessment and Learning for Engineering (TALE), Bali, Indonesia* (pp. 213–218). doi:10.1109/TALE.2013.6654431

Fox, N. J. (2011). Boundary objects, social meanings and the success of new technologies. *Sociology*, *45*(1), 70-85.

Fredricks, J. A., Hofkens, T., Wang, M.-T., Mortenson, E., & Scott, P. (2018). Supporting girls’ and boys’ engagement in math and science learning: A Mixed methods study. *Journal of Research in Science Teaching*, *55*(2), 271-298.

Gerber, B. L., Cavallo, A. M., & Marek, E. A. (2001). Relationships among informal learning environments, teaching procedures and scientific reasoning ability. *International Journal of Science Education*, *23*(5), 535-549.

Hofstein, A., & Rosenfeld, S. (1996). Bridging the gap between formal and informal science learning. *Studies in Science Education*, *28*(1), 87-112.

Hung, D., & Chen, D. T. (2007). Context–process authenticity in learning: Implications for identity enculturation and boundary crossing. *Education Technology and Research Development, 55*, 147-167.

Hwang, G. J., & Tsai, C. C. (2011). Research trends in mobile and ubiquitous learning: A Review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology, 42*(4), E65–E70.

Johanson, G. A., & Brooks, G. P. (2010). Initial scale development: Sample size for pilot studies. *Educational and Psychological Measurement, 70*(3), 394-400.

Johnson, M. L. (1995). Stages and transitions in cognitive development. *Journal of Thought*, *30*(3), 45-69.

Lee, S. C., Irving, K. E. (2018). Development of two-dimensional classroom discourse analysis tool (CDAT): Scientific reasoning and dialog patterns in the secondary science classes. *International Journal of STEM Education, 5*(5). doi:10.1186/s40594-018-0100-0

Lewin, C., & Charania, A. (2018). Bridging formal and informal learning through technology in the twenty-first century: Issues and challenges. In J. Voogt, G. Knezek, R. Christensen, & K.W. Lai (Eds.), *Second Handbook of Information Technology in Primary and Secondary Education, Springer International Handbooks of Education* (pp. 199-215). doi:10.1007/978-3-319-71054-9_13

Looi, C. K., Wong, L. H., So, H., Seow, P., Toh, Y., Norris, C., Soloway, E., Chen, W., & Zhang, B. H. (2009). Anatomy of a mobilized lesson: Learning my way. *Computers & Education, 53*(4), 1120-1132.

Marheineke, M., Habicht, H., & Möslein, K. M. (2016). Bridging knowledge boundaries: The Use of boundary objects in virtual innovation communities. *R & D Management*, *46*(S3), 1084-1098.

Moore, C., & Teather, S. (2013). Engaging students in peer review: Feedback as learning. *Issues in Educational Research, 23*(2), 196-211.

Morgan, D. L. (1997). *Qualitative research methods: Focus groups as qualitative research*. Thousand Oaks, CA: SAGE Publications, Inc. doi:10.4135/9781412984287

Orion, N., Hofstein, A., Tamir, P., & Giddings, G. J. (1997). Development and validation of an instrument for assessing the learning environment of outdoor science activities. *Science Education*, *81*(2), 161-171.

Otero, N., Milrad, M., Rogers, Y., Santos, A. J., Verssimo, M., & Torres, N. (2011). Challenges in designing seamless-learning scenarios: Affective and emotional effects on external representations. *International Journal of Mobile Learning and Organisation*, *5*(1), 15-27.

Patten, B., Arnedillo Sánchez, I., & Tangney, B. (2006). Designing collaborative, constructionist and contextual applications for handheld devices. *Computers & Education, 46*(3), 294–308.

Patrick, P., Mathews, C., & Tunnicliffe, S. D. (2013). Using a field trip inventory to determine if listening to elementary school students' conversations, while on a zoo field trip, enhances preservice teachers' abilities to plan zoo field trips. *International Journal of Science Education*, *35*(15), 2645-2669.

Phelan, P., Davidson, A. L., & Cao, H. T. (1991). Students' multiple worlds: Negotiating the boundaries of family, peer, and school cultures. *Anthropology & Education Quarterly, 22*(3), 224-250.

Pimmer, C. (2016). Mobile learning as boundary crossing: An Alternative route to technology-enhanced learning? *Interactive Learning Environment*, *24*(5), 979-990.

Polman, J. L., & Hope, J. M. G. (2014). Science news stories as boundary objects affecting engagement with science. *Journal of Research in Science Teaching*, *51*(3), 315-341.

Pu, Y-H., Wu, T. T., Chiu, P. S., & Huan, Y-M. (2016). The Design and implementation of authentic learning with mobile technology in vocational nursing practice course. *British Journal of Educational Technology*, *47*(3), 494–509.

Rogers, A. (2014). *The Base of the iceberg. Informal learning and its impact on formal and non-formal learning.* Toronto, Canada: Barbara Budrich Publisher.

Rogoff, B., Callanan, M., Gutiérrez, K., & Erickson, F. (2016). The Organization of informal learning. *Review of Research in Education*, *40*(1), 356-401.

Santos, I. M., & Ali, N. (2012). Exploring the uses of mobile phones to support informal learning. *Education and Information Technologies, 17*, 187–203.

Shadiev, R., Huang, Y.-M., Hwang, W.-Y., & Liu, T.-Y. (2017). Facilitating application of language skills in authentic environments with a mobile learning system. *Journal of Computer Assistant Learning*, *34*(3-4), 42-52.

Stahl, G. (2000). A Model of collaborative knowledge-building. In B. Fishman & S. O'Connor-Divelbiss (Eds.), *Fourth International Conference of the Learning Sciences* (pp. 70-77). Mahwah, NJ: Erlbaum.

Sharples, M., Eileen, S., Shaaron, A., Stamatina, A., Trevor, C., Charles, C., Ann, J., Lucinda, K., Paul, M., & Claire, O. M. (2014). Personal inquiry: Orchestrating science investigations within and beyond the classroom. *Journal of the Learning Sciences*, *24*(2), 308-341. doi:10.1080/10508 406. 2014. 944642

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, translation and boundary objects: Amateurs and professionals in Berkeley's museum of vertebrate zoology, 1907-39. *Social Studies of Science*, *19*(3), 387-420.

Sun, D., & Looi, C.-K. (2019). An Inspiration from border crossing: Principle of boundary activity for integrating learning in the formal and informal spaces. In C.-K. Looi, L.-H. Wong, C. Glahn, & S. Cai. (Eds), *Seamless Learning. Lecture Notes in Educational Technology* (pp. 74-88). Singapore: Springer.

Sun, D., & Looi, C.-K. (2018). Boundary interaction: Towards developing a mobile technology-enabled science curriculum to integrate learning in the informal spaces. *British Journal of Educational Technology, 49*(3), 505-515.

Taber, K. S. (2018). The Use of Cronbach's Alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*, 1273–1296.

Tan, A. L., Jamaludin, A., & Hung, D. (2019). In pursuit of learning in an informal space: A Case study in the Singapore context. *International Journal of Technology and Design Education*. doi:10.1007/s10798-019-09553-1

Tsurusaki, B. K., Barton, A. C., Tan, E., Koch, P. A., & Contento, I. R. (2012). Using transformative boundary objects to create critical engagement in science: A Case study. *Science Education, 97*(1), 1-31.

Wang, C., Fang, T., & Miao, R. (2018). Learning performance and cognitive load in mobile learning: Impact of interaction complexity. *Journal of Computer Assisted Learning*, *34*(6), 917-927.

Wang, Q., Woo, H. L., & Zhao, J. (2009). Investigating critical thinking and knowledge construction in an interactive learning environment. *Interactive Learning Environments*, *17*(1), 95-104.

Wenger, E. (1998). *Communities of practice: Learning, meaning and identity*. Cambridge, UK: Cambridge University Press.

Wertsch, J. V., & Rupert, L. J. (1993). The Authority of cultural tools in a sociocultural approach to mediated agency. *Cognition and Instruction*, *11*(3-4), 227-239.

Wong, L.-H. (2012). A Learner-centric view of mobile seamless learning. *British Journal of Educational Technology*, *43*(1), E19-E23.

Wong, L.-H., Chen, W., & Jan, M. (2012). How artefacts mediate small-group co-creation activities in a mobile-assisted seamless language learning environment? *Journal of Computer Assisted Learning*, *28*(5), 411-424.

Wong, L.-H. (2013). Analysis of students' after-school mobile-assisted artifact creation processes in a seamless language learning environment. *Educational Technology & Society, 16*(2), 198-211.

Zainuddin, Z. (2018). Students' learning performance and perceived motivation in gamified flipped-class instruction. *Computers & Education*, *126*, 75-88.

Zhang, J., Tao, D., Chen, M.-H., Sun, Y., Judson, D., & Naqvi, S. (2018). Co-organizing the collective journey of inquiry with idea thread mapper. *Journal of the Learning Sciences*, *27*(3), 390-430.