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

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

Jul. 2020 Journal of Educational Technology & Society

vol.23
no.3

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 3
July 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.

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 3 2020

Table of contents

Full Length Articles

Park, S., & Weng, W. (2020). The Relationship Between ICT-Related Factors and Student Academic Achievement and the Moderating Effect of Country Economic Indexes Across 39 Countries: Using Multilevel Structural Equation Modelling. *Educational Technology & Society*, *23* (3), 1–15.

The Relationship Between ICT-Related Factors and Student Academic Achievement and the Moderating Effect of Country Economic Indexes Across 39 Countries: Using Multilevel Structural Equation Modelling

Seoyeon Park[1*] and Wenting Weng[2]

[1]Department of Teaching, Learning, and Culture, Texas A&M University, College Station, Texas, USA // [2]Department of Educational Psychology, Texas A&M University, College Station, Texas, USA // pseoyeon5@tamu.edu // vinna_weng@tamu.edu

[*]Corresponding author

(Submitted March 10, 2020 Revised April 14, 2020 Accepted July 8, 2020)

ABSTRACT: This study examined how information and communications technology (ICT) related factors and country-level economic status influence student academic achievement. Two-level structural equation modeling was employed to investigate both student-level and country-level variables, using the PISA 2015 data of ninth-grade students across 39 countries. The findings indicate that: (a) students' interest in ICT, perceived ICT competence, and autonomy had positive impacts on academic performance; (b) GDP per capita had significant interaction effects on the relationship among ICT-related factors (ICT use for studying at school, for entertainment, and perceived ICT autonomy) and academic performance; and (c) a higher level of students' perceived autonomy in ICT resulted in better learning outcomes in countries with less income inequality.

Keywords: Information and Communications Technology (ICT), Multilevel Structural Equation Modeling (MSEM), ICT Competence, ICT Autonomy, Income Inequality

1. Introduction

The integration of information and communications technology (ICT) in teaching and learning has grown in the education field across many countries, not only to promote student achievement but also to enhance equal access to educational platforms and life skills for youths and adults alike (UIS, 2009). It is also believed that the prevalence of ICT in education allows teachers to share their best practices with others, which contributes to the advancement of the overall quality of education (Goktas, Yildirim, & Yildirim, 2009; Murthy, Iyer, & Warriem, 2015). A number of studies have addressed the relationships among ICT-related factors such as frequency of use, availability, interest level, and perceived competence and autonomy in ICT use and academic performance. However, even though the importance and benefits of ICT in education have been widely recognized, no consensus has been reached on how ICT-related factors affect students' academic achievement.

Moreover, most of those studies focused on identifying school- or student-level ICT factors without considering country-level factors (Gómez-Fernández & Mediavilla, 2018), instead of taking a multinational approach. Considering that the national economic status has impacted the country's educational system, including the level of ICT integration in education and students' ICT richness (ITU, 2015), it is crucial to take into account of national-level economic status when probing the impacts of ICT factors on students' achievement Hu, Gong, Lai, and Leung (2018) and Skrybin, Zhang, Liu, and Zhang (2015) analyzed how national level ICT development and students' ICT use influence achievement in reading, math, and science. However, both studies conducted a series of univariate hierarchical linear analyses separately for multivariate outcome variables (math, reading, and science scores), which can inflate the Type I error rate. Moreover, even though the researchers included GDP per capita as a control variable, they did not carefully examine cross-level interactions regarding how a country's economic status can affect the relationships among ICT-related factors and achievement. Inconsistent results on the relationships among ICT-related factors and students' achievement from previous literature can be deeply related to this lack of sufficient and rigid research regarding how national economic status can moderate the overall relationships among ICT-related factors and students' achievement.

Thus, the purpose of this study was to probe the relationship between ICT-related factors and student academic achievement and the moderating effects of country-level economic factors on these relationships. Specifically, (1) each ICT-related factor effect on academic achievement, (2) effect of two country-level economic factors (GDP per capita and the GINI index) on academic achievement, (3) and the cross-level interaction effects will be examined.

2. Theoretical framework

2.1. Student-Level ICT factors and academic achievement

2.1.1. Students' ICT use for entertainment

Students' frequent ICT use for entertainment includes playing games, chatting, or browsing the Internet for fun. Previous studies related to ICT use for entertainment showed contradictory findings regarding its impact on academic achievement. Bulut and Cutumisu (2017) found that ICT use in Finland for entertainment did not benefit academic achievement in mathematics and science. However, several studies found a positive relationship between ICT use for entertainment and student achievement (Gumus & Atalmis, 2011). Other studies reported a nonsignificant relationship between use of ICT for entertainment and student performance (Hu et al., 2018; Juhaňák, Zounek, Záleská, Bárta, & Vlčková, 2018).

2.1.2. Students' interest in ICT use

Student interest in ICT use describes intrinsic motivation toward using ICT. ICT interest is related to positive emotions and enjoyment of using ICT-based products such as mobile devices or computers (Zylka et al., 2015). Interest level is a facet of positive attitudes towards ICT (Hu et al., 2018) that also showed a strong correlation with student engagement in using ICT (Zylka, Christoph, Kroehne, Hartig, & Goldhammer, 2015). Previous studies have had mixed findings regarding the effect of ICT interest on student learning outcomes (Lee & Wu, 2012; Meng, Qiu & Boyd-Wilson, 2019). Some studies showed that ICT interest was positively related to reading literacy (Lee & Wu, 2012; Xiao & Hu, 2019) and mathematics achievement (Meng et al., 2019). However, other studies concluded that ICT interests had no significant impact on mathematics, reading, and science achievement (Juhaňák et al., 2018).

2.1.3. Students' perceived ICT competence

Students' perceived ICT competence refers to their ICT-based knowledge and skills that can be used to perform ICT-related tasks (Meng et al., 2019). There have been conflicting findings regarding the effect of perceived ICT competence on student academic performance. For example, the study by Juhaňák et al. (2018) found that ICT competence had no significant impact on student performance. However, a study by Hu et al. (2018) showed that ICT competence had significant associations with student academic performance in mathematics, reading, and science. Interestingly, Xiao and Hu (2019) found negative associations between reading scores and perceived competence.

2.1.4. Students' perceived ICT autonomy

Students' perceived ICT autonomy can be defined as students taking control of learning via the use of ICT (Fu, 2013). Previous studies showed that students' perceived ICT autonomy had a significantly positive impact on student enjoyment of learning science and interest in science topics (Areepattamannil & Santos, 2019; Hu et al., 2018; Meng et al., 2019). Xiao and Hu (2019) also found a positive relationship between ICT autonomy and reading scores. This might be because students with a high level of autonomy in ICT use can complete learning tasks effectively by doing ICT-related activities such as searching for useful materials (Cárdenas-Claros & Oyanedel, 2016).

2.1.5. ICT academic use at home/school

ICT academic use includes student use of computers or other technologies to do homework or to communicate with friends or teachers regarding schoolwork. The effect of ICT academic use on student achievements is still a matter of debate. Many studies, especially those that employed large-scale data such as the Programme for International Student Assessment (PISA) or the Trends in International Mathematics and Science Study (TIMSS), found a positive relationship between ICT academic use and achievement (Kubiatko & Vlckova, 2010; Delen & Bulut, 2011; Luu & Freeman, 2011). However, some studies found that ICT use for studying had a negative or no significant relationship with academic achievement (Chiao & Chiu, 2018; Park, 2020). Song and

Kang (2012) also suggested that there were trivial but negative relationships between ICT use and achievement in math.

2.2. Country-level economic development factors

This paper explored the moderating effects of two country-level economic development factors (the GINI index and GDP per capita) on the relationship between ICT-related factors and students' academic achievements. The GINI index is a popular measure for income inequality, or the degree to which income is distributed in an unequal manner across a population (Voitchovsky, 2005). It is a measure of statistical dispersion representing the income or wealth distribution in a country (Gastwirth, 1972). It can be calculated by the Lorenz curve framework which produces the comparison of cumulative proportions of the population against cumulative proportions of received income (De Maio, 2007). GDP per capita is regarded as a core indicator of economic performance and used as an aggregate measure of average living standards (OECD, 2009a). In terms of the relationship between GDP per capita and the GINI index, income inequality is known to have a negative impact on economic growth (Cingano, 2014). This relationship indicates that increases in the level of income inequality result in lower transitional GDP per capita growth and have a negative effect on the level of GDP per capita in the long term (Brueckner & Lederman, 2015; Mo, 2000).

2.3. Country-level economic development and student-level ICT factors

Previous research findings have shown that a country's economic development status can affect its ICT access or usage (Aduwa-Ogiegbaen & Iyamu, 2005). ICT access and usage within a country can vary based on the level of economic development because well-developed countries are more likely to better support ICT usage and develop ICT infrastructure in the nation. This can yield higher academic performance because students in wealthy countries may have better ICT-related experience using rich ICT resources that can improve their ICT-related skills (Heinz, 2016).

A country's economic development status influences not only its ICT access or usage but also the country's ICT development level (Skryabin et al., 2015). National ICT development includes readiness of the ICT infrastructure, ICT-trained teachers, ICT support staff, educational software, virtual learning environments, and the impact of these elements on learning (UIS, 2009). A country's economic development status and its national ICT level mutually affect each other; wealthier countries allocate more resources to improve ICT development and overall ICT development plays an important role in promoting the national economy (Luu & Freeman, 2011).

2.4. Country-level economic development factors and students' academic achievements

The effects of country-level economic development status on students' academic achievements have been analyzed in diverse studies in the education field. However, no consensus has emerged on whether or how countries' wealth affects student academic performance (Ensminger, Fothergill, Bornstein, & Bradley, 2003). Some studies have shown that students in high-income countries generally have significantly higher academic achievement than those in low-income countries (Chiu, 2007; Sutton & Soderstrom, 1999). This is because wealthier countries are highly likely to secure a high-quality educational infrastructure including facilities, learning materials, and teachers by increasing spending on education (Saha, 1983; UNICEF, 2001). Other studies suggested that there is a trivial or insignificant relationship between a country's economic status and student academic achievement (Ripple & Luthar, 2000; Seyfried, 1998). This is because the impact of better schools or teachers can vary across countries with different income levels. In developed countries, family background has more of an influence on student academic outcomes than school or teacher factors (Heyneman & Loxley, 1983).

Research concerning the relationship between income inequality and student learning also showed inconsistent results. Income inequality within a country is recognized as one of the strongest factors that can explain differences in student academic achievement. The impact of the income gap on students' academic achievement is more than twice as large as the impact of racial background on academic achievement in the United States (Reardon, 2011). However, cross-national studies regarding the relationship between income inequality and academic outcomes have shown mixed results. Some studies found a modest positive association between income inequality and achievement (Chmielewski & Reardon, 2016). These studies also suggest that income inequality is largely associated with income segregation and child poverty rates, which can affect educational opportunities. On the other hand, several studies, including PISA and TIMSS reports, found that a country's income inequality had a very weak or no effect on student achievement (Dupriez & Dumay, 2006; Marks, 2005; OECD, 2010)

2.5. Rationale for the conceptual framework of the current study

Based on the literature discussed above, ICT-related factors are regarded as having great potential for improving student achievement; however, the effects of these factors on learning have been controversial and need further investigation. Previous research has focused on identifying school- and student-level ICT factors. Few studies consider country-level economic factors that can not only influence school and family ICT richness but also student ICT competence (ITU, 2015). The current study examines the relationship between ICT-related factors and student academic achievement and analyzes the moderating effects of country-level economic factors. The conceptual framework for the study (see Figure 1) includes five ICT factors at the student level: use for study, use for entertainment, interest level, perceived competence, and perceived autonomy. As country-level economic indicators, we measured the country's wealth by using GDP per capita and its economic inequality with the GINI index. Analyzing these student-level ICT factors and country-level economic factors within a model provided us with a comprehensive understanding of how each factor interacted with other factors at the national level and contributed to student learning.

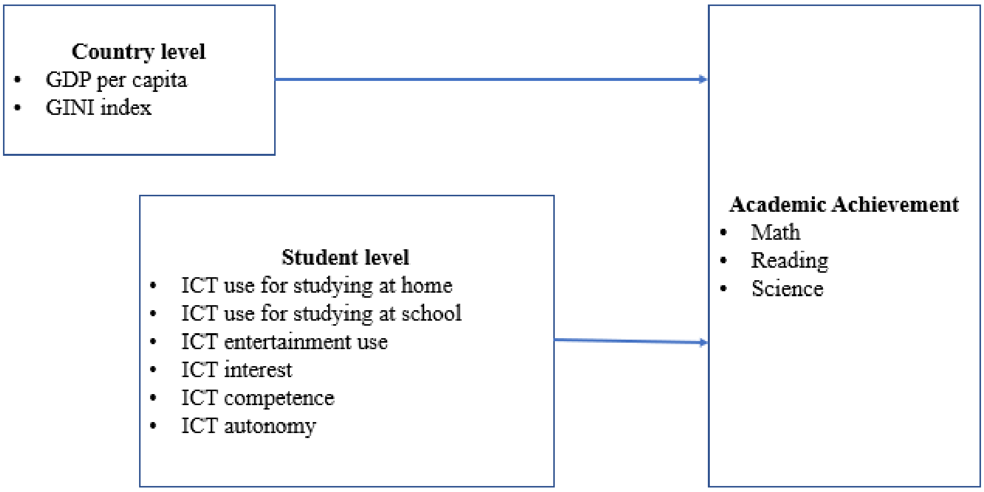


Figure 1. A conceptual framework of the potential influence of ICT factors on academic achievement

Summarizing the purpose of this study and the conceptual framework, we listed two main research questions (RQs) and related hypotheses below:
RQ1: Are student-level ICT variables associated with student academic achievement?

H1a: Higher ICT use at school will be correlated with better academic performance.
H1b: Higher ICT use at home will be correlated with better academic performance.
H1c: Higher students' ICT use for entertainment will be correlated with better academic performance.
H1d: Higher students' interest in ICT use will be correlated with better academic performance.
H1e: Higher students' competence in ICT use will be correlated with better academic performance.
H1f: Higher students' perceived ICT autonomy will be correlated with better academic performance.

RQ2: Does the impact of six ICT-related variables on student academic achievement vary among countries with different levels of economic development?

H2a: The relationship between ICT variables and student academic achievement is stronger in countries having higher GDP.
H2b: The relationship between ICT variables and student academic achievement is stronger in countries having lower GINI.

3. Method

3.1. Data source

This study employed datasets from the OECD's PISA 2015 database and the World Bank. The PISA 2015 dataset assessed 15-year-old students' knowledge and skills in the domains of mathematics, reading, and science across 72 countries and regions. The ICT familiarity questionnaire was optional, so students from some countries completed it. We extracted two country-level indices from World Bank datasets: (1) the national gross domestic product (GDP) per capita and (2) the GINI index representing the income distribution of residents within a nation. Several countries and regions in PISA data were excluded due to a lack of GDP per capita or GINI index information. The excluded countries and regions are Hong Kong, Chinese Taipei, Macao, Spain (regions), Singapore, China (Beijing-Shanghai-Jiangsu-Guangdong), and New Zealand. As a result, a total of 168,098 students from 39 countries were retained in the sample (see Appendix A).

3.2. Instruments

Our study employed student-level and country-level variables following the proposed conceptual framework for examining the relationship between ICT-related factors and student academic achievement. The student-level variables were from the PISA 2015 dataset, while country-level variables were from the World Bank database. The outcome variables evaluated student performance on literacy in mathematics, reading, and science in the PISA 2015 dataset. The PISA 2015 provided 10 plausible values for each student to maximize the estimation of student performance. These plausible values show the distribution of potential scores for students in the population with similar attributes and patterns of item response (OECD, 2009b). According to the PISA data analysis manual (OECD, 2009b), using one plausible value can still provide an unbiased estimate of population parameters. It showed that using one plausible value does not yield a significant difference in the mean estimates or in the standard error estimates with a sample size of 6,400 students (OECD, 2009b). The imputation error of employing one plausible value is also relatively small when the dataset is large and does not make a significant change in the Type I error (OECD, 2009b). Thus, we employed the first plausible value of math, reading, and science for data analysis. Because the scores on math, science, and reading were highly correlated (see Table 1), we created the latent factor "Achievement" using three scores to address multiple outcome variables concurrently.

Student-level ICT-related factors included use or perceptions of studying at school (STUSCH), studying at home (HOMESTU), entertainment (ENTUSE), interest level (INTICT), competence (COMPICT), and autonomy (AUTICT). The related survey items for each of these variables are listed in Appendix B. We generated STUSCH and HOMESTU by averaging the scores of the original survey items, which were only related to the study purpose. The other four ICT variables (i.e., ENTUSE, INTICT, COMPICT, AUTICT) were generated by using weighted likelihood estimates (WLE) (Warm, 1989) and applying a transformation formula (see Figure 2) to the scores of related survey items. All these ICT-related variables in the analysis were group-mean centered by students' countries. At the student level, gender and family socioeconomic status (SES) were included as control variables. The variable of student's family SES was measured by the PISA index of economic, social, and cultural status (ESCS). The scale of ESCS is a transformed WLE score that considers the indicators of parental education, highest parental occupation, and home possessions. The coding of all variables in the multilevel model is presented in Table 2.

$$\delta_f' = \frac{\delta_o - \bar{\delta}_{OECD}}{\sigma_{\delta_{OECD}}}$$

where δ_f' is the final WLE score after transformation, δ_o is the original WLEs in logits, $\bar{\delta}_{OECD}$ is the OECD average of logit scores with weighted country samples, and $\sigma_{\delta_{OECD}}$ is the corresponding standard deviation of the initial WLE score.

Figure 2. The formula of transformation

At the country level, we utilized the logarithm of 2015 GDP per capita and the grand-mean centered 2015 GINI index of each country when creating the GDP and GINI index variables. In addition, we excluded cases with missing data from the analyses. We employed the variance inflation factor (VIF) to check multicollinearity. No VIFs exceeded 10 in the study.

Table 1. Correlations

	Math	Science	Reading	GINI	GDP	STUSCH	HOMESTU	ENTUSE	COMPICT	AUTICT	INTICT
Math	1										
Science	.932	1									
Reading	.865	.919	1								
GINI	-.413	-.324	-.276	1							
GDP	.27	.222	.196	-.467	1						
STUSCH	-.107	-.119	-.141	.015	.051	1					
HOMESTU	-.1	-.105	-.016	.125	-.105	.43	1				
ENTUSE	.015	.008	-.016	-.035	-.002	.31	.415	1			
COMPICT	.083	.098	.081	.014	.028	.187	.215	.374	1		
AUTICT	.149	.16	.112	-.044	.081	.163	.183	.383	.643	1	
INTICT	.044	.085	.063	.039	.049	.17	.195	.377	.527	.443	1

Table 2. Coding of variables

	Scale	Range	Mean	*SD*
Level 1 variables				
STUSCH	1: Never or hardly ever…5: Everyday, centered	[-2.808; 3.812]	0.000	0.941
HOMESTU	1: Never or hardly ever…5: Everyday, centered	[-2.826; 3.576]	0.000	0.998
ENTUSE	1: Never or hardly ever…5: Everyday, WLE	[-4.083; 5.294]	0.000	1.007
INTICT	1: Strongly disagree…4: Strongly agree, WLE	[-3.336; 3.050]	0.000	0.98
COMPICT	1: Strongly disagree…4: Strongly agree, WLE	[-3.047; 2.859]	0.000	0.961
AUTICT	1: Strongly disagree…4: Strongly agree, WLE	[-2.839; 2.456]	0.000	0.98
Gender	1: female, 2: male, standardized	{1; 2}	1.493	-
SES	Index of economic, social and cultural status, WLE	[-6.656; 3.567]	-0.101	1.006
Level 2 variables				
GDP	GDP/c; logarithm	[1.993; 2.446]	2.205	0.105
GINI index	Grand-mean centered	[-9.764; 16.136]	0.000	7.09

3.3. Data analysis

We conducted multilevel structural equation modeling (MSEM) to account for moderation effects of country-level predictors in the relationships among ICT-related factors and achievement. The model was a random-intercepts-random-slopes model (see Figure 3). In the model, the slopes of all ICT-related variables were estimated as random in the within part, which allowed the variables to have different effect sizes across the countries. We labeled the slopes "S_n" between each ICT-related variable and students' achievement. In the between-part of the model, we tested the moderator hypothesis, which was the effects of the level 2 predictors on the slopes "S_n" as well as the mean values of ICT-related variables and student achievement. The multivariate model also included student gender and family social economic status as control variables.

Table 3 shows the changes in intraclass correlation based on different models. In the baseline model, 17.7% of the variation in student science achievement remained unexplained, which can be attributed to the grouping variable (country characteristics). In the within-part of the multilevel model (Level 1) after controlling for gender and SES and adding ICT variables, the unexplained variation was reduced to 12% for science achievement. In the between part of the model (Level 2), this variation was reduced to 11.9%. Likewise, the baseline model showed 23.2% variation in math achievement and 13.6% variation in reading achievement, which was respectively decreased to 16.1% and 8.2% in both the Level 1 and Level 2 models.

Table 3. The results of intraclass correlation

Variable	Baseline Model	Level 1 of the multilevel model	Level 2 of the multilevel model
Math	0.232	0.161	0.161
Science	0.177	0.12	0.119
Reading	0.136	0.082	0.082

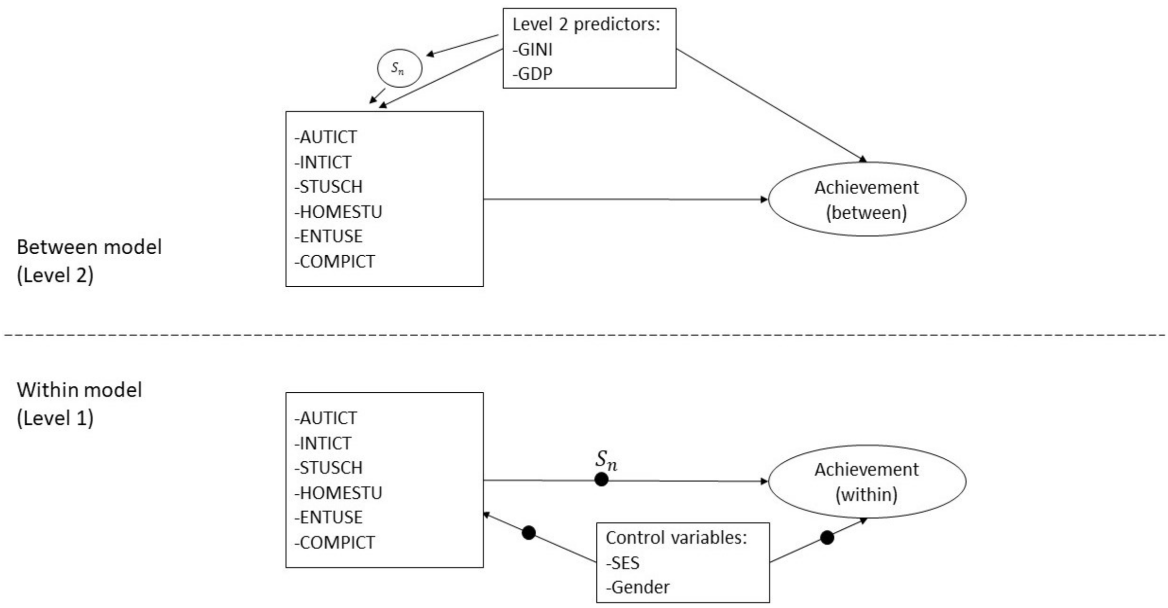


Figure 3. The path paradigm of the MSEM model

4. Results

4.1. Student-level ICT factors

Table 4 reports the estimation result of the multilevel structural equation model using Bayes estimator with Mplus 8.4 (Muthén, 2010). Since the Bayesian 95% credibility intervals (C.I.) of the residual variances do not include zero, all random slopes and intercepts vary across countries even when the country-level predictors are included (Muthén, 2010; Mayerl & Best, 2019). At the within-country level, holding country-level economic factors constant, all ICT-related factors showed significant effects on achievement based on Bayesian 95% credibility intervals. Both STUSCH and HOMESTU negatively correlated with overall achievement at the within-country level. However, both path coefficients (STUSCH = -0.134, HOMESTU = -0.052) were trivial (Cohen, 1992). Students' ICT use for entertainment (ENTUSE) was also negatively associated with achievement, and its effect size was small as well (-0.073). On the other hand, attitudes toward ICT such as students' interest in ICT, perceived competence and autonomy in ICT showed significantly positive relationships with academic achievement. Those variables regarding ICT attitudes showed positive small effects on academic achievement. The path coefficients were 0.045 (INTICT), 0.022 (COMPICT), 0.122 (AUTICT) respectively. Overall, we found that ICT related variables have small but significant impacts on achievement within-country level. However, since there are significant interaction effects of GDP per capita or GINI index, the effects of STUSCH, ENTUSE, and AUTICT on achievement are conditional and the interpretation of these random intercepts depend on the value of the country's GDP per capita or GINI index.

4.2. Country-level economic factors as moderators

While GDP per capita had no significant effect on academic achievement (p = .225, Bayesian C.I. includes zero), the GINI index had a significant negative path coefficient (-0.440, p < .001). Thus, we found that the income inequality of a country had a medium negative effect on achievement.

In terms of cross-level interactions, GDP per capita exhibited a significant interaction effect on the relationship between achievement and students' ICT use for studying at school. The path coefficient was 0.328, implying that a country's national wealth has a moderately positive effect on the relationship between students' achievement and their ICT use for studying at school. GDP per capita also had a moderately positive interaction effect (0.251, p = .027) on the relationship between achievement and their ICT autonomy. However, GDP per capita showed a moderately negative interaction effect (-0.473, p < .001) on the relationship between achievement and their ICT use for entertainment. The GINI index showed a moderately negative interaction effect (-0.417, p < .001) on the relationship between achievement and perceived ICT autonomy, but a moderately positive interaction effect (0.327, p < .001) on the relationship between achievement and ICT use for entertainment.

Table 4. Multilevel SEM Results

	Within-Country Model		95% C.I.	
	Standardized Estimate	Posterior *SD*	Lower 2.5%	Upper 2.5%
Achievement by				
Math	0.936^{***}	0.000	0.936	0.937
Reading	0.920^{***}	0.000	0.936	0.937
Science	0.990^{***}	0.000	0.990	0.990
Level 1 predictors				
STUSCH → Achievement	-0.134^{***}	0.002	-0.139	-0.129
HOMESTU → Achievement	-0.052^{***}	0.003	-0.057	-0.047
ENTUSE → Achievement	-0.073^{***}	0.003	-0.078	-0.067
INTICT → Achievement	0.045^{***}	0.003	0.040	0.050
COMPICT → Achievement	0.022^{***}	0.003	0.016	0.029
AUTICT → Achievement	0.122^{***}	0.003	0.115	0.128
Gender → Achievement	0.028^{***}	0.002	0.024	0.032
SES → Achievement	0.359^{***}	0.002	0.355	0.363
Residual Variances				
MATH	0.123^{***}	0.001	0.122	0.125
READ	0.153^{***}	0.001	0.151	0.154
SCIENCE	0.020^{***}	0.000	0.019	0.021
Achievement	0.795^{***}	0.002	0.792	0.798
	Between-Country Model		95% C.I.	
	Standardized Estimate	Posterior *SD*	Lower 2.5%	Upper 2.5%
Achievement by				
Math	0.926^{***}	0.362	0.024	0.125
Reading	0.949^{***}	0.023	-0.998	-0.914
Science	0.985^{***}	0.016	-0.999	-0.921
Level 2 predictors				
Gini → Achievement	-0.440^{***}	0.139	-0.598	-0.067
GDP → Achievement	-0.082	0.103	-0.280	0.115
Cross-level interactions				
Gini*STUSCH → Achievement	-0.016	0.119	-0.245	0.218
GDP*STUSCH → Achievement	0.328^{***}	0.110	0.084	0.516
Gini*HOMESTU → Achievement	0.060	0.120	-0.181	0.291
GDP*HOMESTU → Achievement	-0.032	0.132	-0.293	0.222
Gini*ENTUSE → Achievement	0.327^{***}	0.108	0.093	0.514
GDP*ENTUSE → Achievement	-0.473^{***}	0.105	-0.642	-0.238
Gini*INTICT → Achievement	0.189	0.117	-0.054	0.404
GDP*INTICT → Achievement	0.064	0.124	-0.190	0.292
Gini*COMPICT → Achievement	0.150	0.121	-0.096	0.374
GDP*COMPICT → Achievement	-0.129	0.129	-0.366	0.130
Gini*AUTICT → Achievement	-0.417^{***}	0.103	-0.590	-0.192
GDP*AUTICT → Achievement	0.251	0.122	-0.005	0.474
Residual Variances				
MATH	0.142^{***}	0.356	0.065	1.000
READ	0.066^{***}	0.042	0.004	0.165
SCIENCE	0.051^{***}	0.040	0.003	0.151
Achievement	0.735^{***}	0.083	0.567	0.885
STUSCH	0.851^{***}	0.094	0.632	0.984
HOMESTU	0.972^{***}	0.041	0.846	0.999
ENTUSE	0.753^{***}	0.086	0.577	0.907
INTICT	0.934^{***}	0.069	0.747	0.997
COMPICT	0.953^{***}	0.046	0.827	0.998
AUTICT	0.835^{***}	0.069	0.683	0.951

Note. $^{***}p < .001$ (one-tailed p-value based on the posterior distribution). The *p*-value is the proportion of the posterior distribution that is below zero for a positive estimate.

5. Discussion

5.1. Student-level ICT factors and academic achievement

5.1.1. Students' ICT use for entertainment

At the within-country level, students' ICT use for entertainment showed negative associations with achievement. This result was consistent with those of other studies such as Bulut and Cutumisu (2017), Petko, Cantieni, and Prasse (2017). Following the result of this study and supportive results from other literature, students' ICT use for entertainment seems to negatively affect students' academic performances. Considering that students' ICT use for entertainment includes playing games, browsing the Internet for fun, and downloading and enjoying movies/music, ICT use for leisure can distract students from learning activities or schoolwork. As ICT advances, students can more easily access a variety of materials for fun by using ICT. Thus, the possibility that students' learning can be harmed by ICT entertainment use also tends to be higher than before.

5.1.2. Students' interest in ICT use

Students' interest in ICT exhibited significant positive relationships with achievement. This result supports other studies' results that interest in ICT has a positive relationship with academic achievement (Hu et al., 2018; Scherer, Rohatgi, & Hatlevik, 2017). Some of the studies showed that ICT interest had no significant impact on student's academic performance (Juhaňák et al., 2018), but most of those studies only analyzed a specific country. According to the result of this study and other literature, students with higher interests in ICT use were highly likely to have better learning outcomes. This can be explained as students with a higher interest in ICT will engage in learning activities using computers or the Internet more often than other students (Lee & Wu, 2012; Scherer et al., 2017). Moreover, these students would be more motivated and have more positive attitudes toward learning with technology.

5.1.3. Students' perceived ICT competence

Students' perceived ICT competence also showed a significant positive impact on students' academic performance. This result is also consistent with other studies such as Hosein et al. (2010), Hu et al. (2018), or Selwyn and Husen (2010). According to these findings, students with higher perceived competence in ICT were more likely to frequently use software or online resources for studying than those with lower ICT competence (Hosein et al., 2010). However, some studies claimed that ICT competence had no or negative relationship with achievement (Juhaňák et al., 2018; Xiao & Hu, 2019). The results can differ depending on students' grade levels and what country the study analyzed. Students' perceived ICT competence in their academic achievement tends to diminish as the academic year progresses because students become more competent as they get older (Selwyn & Husen, 2010).

5.1.4. Students' perceived ICT autonomy

The result showed that students' perceived ICT autonomy had a significant positive effect on students' academic performance, and its effect size is larger than ICT interest or competence. This result supports other literature that ICT autonomy had a significantly positive relationship with academic achievement (Cárdenas-Claros & Oyanedel, 2016; Xiao & Hu, 2019; Meng et al., 2019). Students with higher autonomy in ICT use have a tendency to take more control of their learning process with technology. Autonomy implicates that students can regulate their learning and can use ICT toward completing tasks and achieve mastery (Fu, 2013). Thus, ICT autonomy can have a more powerful impact on students' learning outcomes than other attitudes regarding ICT use, which are not always related to students' learning activities. Furthermore, once students recognize that they can control their learning with ICT, then they can strengthen their autonomy in ICT use and obtain more knowledge by using ICT effectively (Serhan, 2009).

5.1.5. ICT academic use at home/school

Students' ICT academic use at home and school showed negative relationships with their academic achievement. This result is in agreement with previous studies that ICT academic use had a negative effect on students' academic performances (Song & Kang, 2012; Chiao & Chiu, 2018; Park, 2020). Some studies found that there might be a positive relationship between students' academic use and achievement (Luu & Freeman, 2011; Delen & Bulut, 2011). However, Luu and Freeman (2011) analyzed only Canada and Australia, and Delen and Bulut's work (2011) also solely used Turkey data. Thus, we can find that students' ICT academic use is likely to decrease achievement across countries according to literature. One possible explanation of why it showed a negative impact on achievement would be regarding the PISA ICT questionnaire. Questions that are related to students' academic use only ask the frequency of using ICT. The frequent ICT use for doing homework or searching for studying resources cannot guarantee the quality of students' ICT use for learning. The frequency of ICT academic use cannot provide us with information concerning how much students can focus on tasks while using ICT for studying or how much time they spend studying with the use of ICT.

5.2. Moderating effects of country-level economic factors on the relationship between ICT factors and academic achievement

In terms of GDP per capita, it had no significance on students' academic achievement, which aligns with previous research findings (e.g., Ripple & Luthar, 2000). Even though the wealthier countries may have better ICT infrastructure for education (UNICEF, 2001), the impact of a country's economic size turned out to be insignificant. We also found that the influence of students' ICT use for studying at school and perceived ICT autonomy on achievement was stronger in countries with higher GDP per capita. This supports the idea that a country's economic development levels are associated with student achievement as it affects school-level ICT resources and skills such as infrastructure, ICT support staff, and educational software. Furthermore, ICT richness in school and classroom environments allow students to have more opportunities to complete schoolwork independently and to reflect on their learning progress using technology-mediated communication.

Our findings showed that students in a country with less income inequality would have better academic achievement than those in a country with more income inequality. Therefore, income inequality can lead to a significant academic achievement gap (Reardon, 2011). Income inequality showed positive interaction effects on the impacts of students' perceived ICT competence and ICT interest on academic performance even though these factors were not statistically significant. This reveals that students' perceived attitudes toward ICT can more greatly impact academic performance in countries with higher income inequality than other countries. This shows that students who can exert a higher level of autonomy in ICT use are likely to get better learning outcomes in countries with less income inequality based on the negative and significant effects of a country's income inequality on the relationship between students' perceived ICT autonomy and academic achievement.

Overall, we found that income inequality has broader and more extensive impact on the relationship between ICT-related attitudes and students' achievement than a country's wealth itself. This suggests that students from lower-income families in these countries are highly likely to lag behind academically due to a lack of ICT or resource availability. Indeed, Rideout and Katz (2016) found that around 40% of parents without computers or Internet access cannot afford to provide their children necessary resources. Several studies have also reported that funding is the primary obstacle creating a digital divide, which is a gap between students who can utilize technology to acquire knowledge and those who cannot (Rideout and Katz, 2016). Hence, to solve this educational inequality based on addressing income inequality and the digital divide, it is essential to promote equitable access to digital technologies through related policies, including discounted internet access and expanding ICT infrastructure in public areas such as schools or libraries for low-income families (Kelley-Salinas, 2000). To support the ICT competence, interest, and autonomy of students from low-income families, supporting a variety of education programs that can improve students' ICT attitudes and skills that can lead to improving their learning is important (Koh, Chai, & Lim, 2017). This can also be obtained by improving teachers' ICT skills through professional development opportunities and continuous support from government, district, and school communities (Akbaba-Altun, 2006; Jung, 2005).

6. Conclusion

This study contributes to the education field via multilevel analysis of the relationship between student-level ICT factors and academic achievement and of the moderating effect of national-level economic indices on these relationships. The results indicate that students' use of ICT for both studying and entertainment had a negative association with their academic achievement; however, interest, perceived competence, and autonomy in ICT use showed a positive impact on students' learning. We employed two national-level economic indices, GDP per capita and the Gini index, to analyze how a country's economic status can moderate the relationships between

ICT-related factors and achievement. GDP per capita showed significant interaction effects on the relationship between achievement and students' ICT use for studying at school, entertainment, and perceived ICT autonomy. The GINI index also showed medium interaction effects on relationships between achievement and ICT variables (ICT use for entertainment and students' perceived ICT autonomy). This suggests that the effects of ICT-related factors on achievement should be interpreted carefully in the context of national-level wealth and income inequality.

This study has a few limitations. First, we analyzed the responses of students from countries that participated in the PISA ICT questionnaire. Because not all countries responded to the ICT questionnaire, the results of this paper are limited to those selected countries. Second, the PISA ICT questionnaire consists of questions on frequency or the availability of ICT use, which cannot capture the quality of ICT use. For future studies, more factors related to the quality of student ICT use should be explored to analyze their influences on students' learning outcomes. Variables at other levels (e.g. teacher or school) that might influence the relationship between ICT-related factors on student academic performance should also be developed.

References

Aduwa-Ogiegbaen, S. E., & Iyamu, E. O. S. (2005). Using information and communication technology in secondary schools in Nigeria: Problems and prospects. *Educational Technology & Society, 8*(1), 104-112.

Akbaba-Altun, S. (2006). Complexity of integrating computer technologies into education in Turkey. *Educational Technology & Society, 9*(1), 176-187.

Areepattamannil, S., & Santos, I. M. (2019). Adolescent students' perceived information and communication technology (ICT) competence and autonomy: Examining links to dispositions toward science in 42 countries. *Computers in Human Behavior, 98,* 50-58. doi:10.1016/j.chb.2019.04.005

Brueckner, M., & Lederman, D. (2015). *Effects of income inequality on aggregate output*. Washington, DC: World Bank.

Bulut, O., & Cutumisu, M. (2017). When technology does not add up: ICT use negatively predicts Mathematics and Science achievement for Finnish and Turkish students in PISA 2012. In *EdMedia+ Innovate Learning* (pp. 935-945). Washington, DC: Association for the Advancement of Computing in Education (AACE).

Cárdenas-Claros, M., & Oyanedel, M. (2016). Teachers' implicit theories and use of ICTs in the language classroom. *Technology, Pedagogy and Education, 25(*2), 207-225. doi:10.1080/1475939X.2014.988745

Chiao, C., & Chiu, C. H. (2018). The Mediating effect of ICT usage on the relationship between students' socioeconomic status and achievement. *The Asia-Pacific Education Researcher, 27*(2), 109-121.

Chiu, M. M. (2007). Families, economies, cultures, and science achievement in 41 countries: Country-, school-, and student-level analyses. *Journal of Family Psychology, 21*(3), 510-519. doi:10.1037/0893-3200.21.3.510

Chmielewski, A. K., & Reardon, S. F. (2016). Patterns of cross-national variation in the association between income and academic achievement. *AERA Open, 2*(3), 2332858416649593. doi:10.1177/2332858416649593

Cingano, F. (2014). Trends in income inequality and its impact on economic growth. In *OECD Social, Employment and Migration Working Papers* (No. 163). Paris, France: OECD Publishing. doi:10.1787/5jxrjncwxv6j-en

Cohen, J. (1992). A Power primer. *Psychological Bulletin, 112*, 155–159.

Delen, E., & Bulut, O. (2011). The Relationship between students' exposure to technology and their achievement in science and math. *The Turkish Online Journal of Educational Technology, 10*(3), 311-317.

De Maio, F. G. (2007). Income inequality measures. *Journal of Epidemiology & Community Health, 61*(10), 849-852. doi:10.1136/jech.2006.052969

Dupriez, V., & Dumay, X. (2006). Inequalities in school systems: Effect of school structure or of society structure? *Comparative Education, 42*(2), 243–260. doi:10.1080/03050060600628074

Ensminger, M. E., Fothergill, K. E., Bornstein, M. H., & Bradley, R. H. (2003). A Decade of measuring SES: What it tells us and where to go from here. In M. H. Bornstein & R. H. Bradley (Eds.), *Socioeconomic status, parenting, and child development* (pp. 13-27). Mahwah, NJ: Erlbaum.

Fu, J. (2013). Complexity of ICT in education: A Critical literature review and its implications. *International Journal of education and Development using ICT, 9*(1), 112-125.

Gastwirth, J. L. (1972). The Estimation of the Lorenz curve and Gini index. *The Review of Economics and Statistics, 54*(3), 306-316.

Goktas, Y., Yildirim, S., & Yildirim, Z. (2009). Main barriers and possible enablers of ICTs integration into pre-service teacher education programs. *Educational Technology & Society, 12*(1), 193-204.

Gómez-Fernández, N., & Mediavilla, M. (2018). Do information and communication technologies (ICT) improve educational outcomes? Evidence for Spain in PISA 2015. *IEB Working Paper*. doi:10.2139/ssrn.3290513

Gumus, S., & Atalmis, E. H. (2011). Exploring the relationship between purpose of computer usage and reading skills of Turkish students: evidence from PISA 2006. *Turkish Online Journal of Educational Technology-TOJET, 10*(3), 129-140.

Heinz, J. (2016). Digital skills and the influence of students' socio-economic background. An Exploratory study in German elementary schools. *Italian Journal of Sociology of Education, 8*(2), 186-212. doi:10.14658/pupj-ijse-2016-2-9

Heyneman, S. P., & Loxley, W. A. (1983). The Effect of primary-school quality on academic achievement across twenty-nine high-and low-income countries. *American Journal of Sociology, 88*(6), 1162–1194.

Hosein, A., Ramanau, R., & Jones, C. (2010). Learning and living technologies: A Longitudinal study of first-year students' frequency and competence in the use of ICT. *Learning, media and technology, 35*(4), 403-418. doi:10.1080/17439884.2010.529913

Hu, X., Gong, Y., Lai, C., & Leung, F. K. (2018). The Relationship between ICT and student literacy in mathematics, reading, and science across 44 countries: A Multilevel analysis. *Computers & Education, 125*, 1-13. doi:10.1016/j.compedu.2018.05.021

International Telecommunication Union (ITU). (2015). *Measuring the information society report 2015 - Executive summary*. Retrieved from https://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2015/MISR2015-ES-E.pdf

Juhaňák, L., Zounek, J., Záleská, K., Bárta, O., & Vlčková, K. (2018). The Relationship between sudents' ICT use and their school performance: Evidence from PISA 2015 in the Czech Republic. *Orbis scholae, 12*(2), 37-64. doi:10.14712/23363177.2018.292

Jung, I. (2005). ICT-pedagogy integration in teacher training: Application cases worldwide. *Educational Technology & Society, 8*(2), 94-101.

Kelley-Salinas, G. (2000). Different educational inequalities: ICT an option to close the gaps. In *Schooling for tomorrow: Learning to bridge the digital divide* (pp. 21-36). Paris, France: OECD.

Koh, J. H. L., Chai, C. S., & Lim, W. Y. (2017). Teacher professional development for TPACK-21CL: Effects on teacher ICT integration and student outcomes. *Journal of Educational Computing Research, 55*(2), 172-196. doi:10.1177/0735633116656848

Kubiatko, M., & Vlckova, K. (2010). The Relationship between ICT use and science knowledge for Czech students: A Secondary analysis of PISA 2006. *International Journal of Science and Mathematics Education, 8*(3), 523-543.

Lee, Y. H., & Wu, J. Y. (2012). The Effect of individual differences in the inner and outer states of ICT on engagement in online reading activities and PISA 2009 reading literacy: Exploring the relationship between the old and new reading literacy. *Learning and Individual Differences, 22*(3), 336-342. doi:10.1016/j.lindif.2012.01.007

Luu, K., & Freeman, J. G. (2011). An Analysis of the relationship between information and communication technology (ICT) and scientific literacy in Canada and Australia. *Computers & Education, 56*(4), 1072-1082. doi:10.1016/j.compedu.2010.11.008

Marks, G. N. (2005). Cross-national differences and accounting for social class inequalities in education. *International Sociology, 20*(4), 483–505. doi:10.1177/0268580905058328

Mayerl, J., & Best, H. (2019). Attitudes and behavioral intentions to protect the environment: How consistent is the structure of environmental concern in cross-national comparison? *International Journal of Sociology, 49*(1), 27-52. doi:10.1080/00207659.2018.1560980

Meng, L., Qiu, C., & Boyd-Wilson, B. (2019). Measurement invariance of the ICT engagement construct and its association with students' performance in China and Germany: Evidence from PISA 2015 data. *British Journal of Educational Technology, 50*(6), 3233-3251. doi:10.1111/bjet.12729

Mo, P. H. (2000). Income inequality and economic growth. *Kyklos, 53*(3), 293-315.

Murthy, S., Iyer, S., & Warriem, J. (2015). ET4ET: A Large-scale faculty professional development program on effective integration of educational technology. *Educational Technology & Society, 18*(3), 16-28.

Muthén, B. (2010). *Bayesian analysis in Mplus: A Brief introduction.* Retrieved from https://www.statmodel.com/download/IntroBayesVersion%203.pdf

OECD (2009a). *National Accounts of OECD Countries 2009, Volume I, Main Aggregates.* OECD Publishing. doi:10.1787/na_vol_1-2009-en-fr

OECD. (2009b). *PISA Data Analysis Manual SPSS*. Paris, France: Author.

OECD. (2010). *PISA 2009 Results: Overcoming Social Background – Equity in Learning Opportunities and Outcomes (Volume II).* OECD Publishing. doi:10.1787/9789264091504-en

Park, S. (2020, April). *ICT usage and perceived control as mediators in the relationship between SES and Math achievement*. Paper presented at the Annual meeting of the American Educational Research Association, San Francisco, California.

Petko, D., Cantieni, A., & Prasse, D. (2017). Perceived quality of educational technology matters: A Secondary analysis of students' ICT use, ICT-related attitudes, and PISA 2012 test scores. *Journal of Educational Computing Research*, *54*(8), 1070-1091. doi:10.1177/0735633116649373

Reardon, S. F. (2011). The Widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In G. J. Duncan, & R. J. Murnane (Eds.), *Whither opportunity?* (pp. 91-116). New York, NY: Russell Sage Foundation.

Rideout, V. J., & Katz, V. S. (2016). *Opportunity for all? Technology and learning in lower-income families. A Report of the Families and Media Project.* New York, NY: The Joan Ganz Cooney Center at Sesame Workshop.

Ripple, C. H., & Luthar, S. S. (2000). Academic risk among inner-city adolescents: The Role of personal attributes. *Journal of School Psychology, 38*(3), 277–298. doi:10.1016/S0022-4405(00)00032-7

Saha, L. J. (1983). Social structure and teacher effects on academic achievement: A Comparative analysis. *Comparative Education Review*, *27*(1), 69-88.

Scherer, R., Rohatgi, A., & Hatlevik, O. E. (2017). Students' profiles of ICT use: Identification, determinants, and relations to achievement in a computer and information literacy test. *Computers in Human Behavior*, *70*, 486-499. doi:10.1016/j.chb.2017.01.034

Selwyn, N., & Husen, O. (2010). The Educational benefits of technological competence: An Investigation of students' perceptions. *Evaluation & Research in Education*, *23*(2), 137-141. doi:10.1080/09500790.2010.483515

Serhan, D. (2009). Preparing pre-service teachers for computer technology integration. *International journal of instructional media*, *36*(4), 439-448.

Seyfried, S. F. (1998). Academic achievement of African American preadolescents: The Influence of teacher perceptions. *American Journal of Community Psychology, 26*(3), 381–402.

Skryabin, M., Zhang, J., Liu, L., & Zhang, D. (2015). How the ICT development level and usage influence student achievement in reading, mathematics, and science. *Computers & Education*, *85*, 49-58. doi:10.1016/j.compedu.2015.02.004

Song, H. D., & Kang, T. (2012). Evaluating the impacts of ICT use: A Multi-level analysis with hierarchical linear modeling. *Turkish Online Journal of Educational Technology*, *11*(4), 132-140.

Sutton, A., & Soderstrom, I. (1999). Predicting elementary and secondary school achievement with school-related and demographic factors. *Journal of Educational Research, 92*(6), 330–338. doi:10.1080/00220679909597616

The UNESCO Institute of Statistics (UIS). (2009). *Guide to measuring information and communication technologies (ICT) in education*. Montreal, Canada: UNESCO Institute for Statistics.

The United Nations Children's Fund (UCEF). (2001). *The State of the world's children 2001.* New York, NY: UNICEF.

Voitchovsky, S. (2005). Does the profile of income inequality matter for economic growth? *Journal of Economic growth*, *10*(3), 273-296.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54*(3), 427-450.

Xiao, Y., & Hu, J. (2019). Regression analysis of ICT impact factors on early adolescents' reading proficiency in five high-performing countries. *Frontiers in psychology*, *10,* 1-14. doi:10.3389/fpsyg.2019.01646

Zylka, J., Christoph, G., Kroehne, U., Hartig, J., & Goldhammer, F. (2015). Moving beyond cognitive elements of ICT literacy: First evidence on the structure of ICT engagement. *Computers in Human Behavior*, *53*, 149-160. doi:10.1016/j.chb.2015.07.008

Appendix A: Country and sample of the study

Country	Sample size
Australia	8,298
Austria	4,346
Belgium	5,882
Bulgaria	3,076
Brazil	6,847
Switzerland	3,619
Chile	4,597
Colombia	6,833
Costa Rica	3,594
Czech Republic	4,617
Denmark	4,385
Dominican Republic	1,996
Spain	4,679
Estonia	4,023
Finland	4,273
France	3,765
United Kingdom	3,665
Greece	3,654
Croatia	3,860
Hungary	3,681
Ireland	3,942
Iceland	2,326
Israel	3,990
Italy	7,609
Japan	5,085
Korea	4,645
Lithuania	4,316
Luxembourg	3,079
Latvia	3,497
Mexico	5,194
Netherlands	4,115
Peru	4,070
Poland	3,378
Portugal	5,204
Russia	3,838
Slovakia	4,030
Slovenia	4,068
Sweden	3,299
Uruguay	2,723

Appendix B: Description of ICT variables and survey items

Variable	Item code	Survey item
STUSCH	IC011Q03TA	Browsing the Internet for schoolwork.
	IC011Q07TA	Practicing and drilling, such as for foreign language learning or mathematics.
	IC011Q08TA	Doing homework on a school computer.
	IC011Q09TA	Using school computers for group work and communication with other students.
HOMESTU	IC010Q01TA	Browsing the Internet for schoolwork (e.g., for preparing an essay or presentation).
	IC010Q02TA	Browsing the Internet to follow up lessons, e.g., for finding explanations.
	IC010Q03TA	Using email for communication with other students about schoolwork.
	IC010Q04TA	Using email for communication with teachers and the submission of homework or other schoolwork.
	IC010Q05NA	Using social networks for communication with other students about schoolwork (e.g., <Facebook>, <MySpace>).
	IC010Q09NA	Doing homework on a computer.
	IC010Q10NA	Doing homework on a mobile device.
	IC010Q11NA	Downloading learning apps on a mobile device.
		How often do you use digital devices for the following activities outside of school?
ENTUSE	IC008Q01TA	Playing one-player games.
	IC008Q02TA	Playing collaborative online games.
	IC008Q03TA	Using email.
	IC008Q04TA	<Chatting online> (e.g., <MSN®>).
	IC008Q05TA	Participating in social networks (e.g., <Facebook>, <MySpace>).
	IC008Q07NA	Playing online games via social networks (e.g., <Farmville®>, <The Sims Social>).
	IC008Q08TA	Browsing the Internet for fun (such as watching videos, e.g., <YouTube>.
	IC008Q09TA	Reading news on the Internet (e.g., current affairs).
	IC008Q10TA	Obtaining practical information from the Internet (e.g., locations, dates of events).
	IC008Q11TA	Downloading music, films, games or software from the internet.
	IC008Q12TA	Uploading your own created contents for sharing (e.g., music, poetry, videos, computer programs).
	IC008Q13NA	Downloading new apps on a mobile device.
		Thinking about your experience with digital media and digital devices: to what extent do you disagree or agree with the following statements?
INTICT	IC013Q01NA	I forget about time when I'm using digital devices.
	IC013Q04NA	The Internet is a great resource for obtaining information I am interested in (e.g., news, sports, dictionary).
	IC013Q05NA	It is very useful to have social networks on the Internet.
	IC013Q11NA	I am really excited discovering new digital devices or applications.
	IC013Q12NA	I really feel bad if no internet connection is possible.
	IC013Q13NA	I like using digital devices.
		Thinking about your experience with digital media and digital devices: to what extent do you disagree or agree with the following statements?
COMPICT	IC014Q03NA	I feel comfortable using digital devices that I am less familiar with.
	IC014Q04NA	If my friends and relatives want to buy new digital devices or applications, I can give them advice.
	IC014Q06NA	I feel comfortable using my digital devices at home.
	IC014Q08NA	When I come across problems with digital devices, I think I can solve them.
	IC014Q09NA	If my friends and relatives have a problem with digital devices, I can help them.
		Thinking about your experience with digital media and digital devices: to what extent do you disagree or agree with the following statements?
AUTICT	IC015Q02NA	If I need new software, I install it by myself.
	IC015Q03NA	I read information about digital devices to be independent.
	IC015Q05NA	I use digital devices as I want to use them.
	IC015Q07NA	If I have a problem with digital devices I start to solve it on my own.
	IC015Q09NA	If I need a new application, I choose it by myself.

Yang, X., Lin, L., Wen, Y., Cheng, P.-Y., Yang, X., & An, Y. (2020). Time-Compressed Audio on Attention, Meditation, Cognitive Load, and Learning. *Educational Technology & Society*, *23* (3), 16–26.

Time-Compressed Audio on Attention, Meditation, Cognitive Load, and Learning

Xiaozhe Yang[1], Lin Lin[2*], Yi Wen[1], Pei-Yu Cheng[3], Xue Yang[2] and Yunjo An[2]

[1]Institute of Curriculum and Instruction, East China Normal University, China // [2]Department of Learning Technologies, University of North Texas, USA // [3]Department of Engineering Science, National Cheng Kung University, Taiwan // worldetyang@gmail.com // Lin.Lin@unt.edu // wenyiyimy@163.com // peiyu.cheng.tw@gmail.com // sketchmichelle@gmail.com // Yunjo.An@unt.edu

[*]Corresponding author

(Submitted May 9, 2020; Revised June 24, 2020; Accepted July 16, 2020)

ABSTRACT: This study examined how three auditory lectures delivered at different speeds – normal (1.0x), fast (1.5x) and very fast (3.0x) speeds – affected the graduate students' attention, cognitive load, and learning that were assessed by pre- and post-comprehension tests, cognitive-load questionnaire, and Electroencephalography (EEG) device. The results showed that there was no significant difference in the students' attention, cognitive load, and learning performance between the normal (1.0x) and 1.5x speed. However, when the auditory lecture speed reached three times of its original speed (3.0x), the students' comprehension scores were significantly lower both in the immediate and (one-week) delayed recall tests, than those in the other two speed conditions. When listening to the lecture at the 3.0x speed, the learners had a higher level of attention and cognitive load. The study provided insights for teaching, instructional design, and learning.

Keywords: Time-compression, Audible, Attention, Cognitive Load, Electroencephalography (EEG)

1. Introduction

New multimedia technologies have made auditory and visual learning more popular than ever (Wang, Wu, & Wang, 2009). An increasing number of young people prefer to listen to or watch videos rather than to read books when seeking information or learning new things (Evans, 2008). Audible.com, one of the world's largest producers of downloadable audiobooks, sells digital audiobooks, radio and TV programs, and audio versions of magazines and newspapers of all kinds. Audiobooks are valuable alternatives to music and podcasts. With the rising popularity of Audible and other audiobook providers, it is easier than ever to stimulate one's mind listening to news and stories while doing other things. The ubiquitous online learning has also facilitated auditory and visual learning opportunities with audio-video lectures in formal and online learning environments (Kress & Selander, 2012).

However, auditory narratives present certain constraints. For instance, it may take longer for a learner to listen to or watch someone present information than to read the texts for the same information (Barron & Kysilka, 1993; Koroghlanian & Sullivan, 2000). Studies have shown that adults in most English-speaking countries can read 280 words per minute, while the normal speed of speaking is only 120-180 words per minute (Pastore, 2012). The native Chinese speakers can usually read at an average speed of 295±51 words per minute (Wang et al., 2018), while a survey of broadcasters showed that each word spoken in Mandarin Chinese would take about 0.224s (Lee & Chan, 2003). That is, the average speech rate of the Mandarin speakers is only 260-300 words per minute. In addition, reading allows a reader to adjust the speed him or herself, while the speed of the auditory narratives is highly dependent on the timing of the auditor (Orr, Friedman, & Graae, 1969). This inflexibility may be in conflicts with the desire of self-directed learners to increase learning efficiency and effectiveness (Broadbent, 2017). In fact, different teaching approaches and learning strategies are constantly adapted to increase learning effectiveness and efficiencies (i.e., achieving the best learning with the least amount of time).

Time compression is a technique to increase the speed of auditory lectures without distorting the tones, intonations, or the output quality of the spoken lectures (Barabasz, 1968; Goldhaber, 1970). Researchers have begun to examine the impact of time compression techniques on cognition and learning. When the time is compressed by 50% or 1.5x speed, it means that the learning task can be completed in half of the time, which is very appealing to the learners. Some researchers found that time compression was directly proportional to the degree of hearing difficulties, and as the compression rate increased, cognitive difficulties began to increase (King & Behnke, 1989). Yet, some other studies showed that there was no difference in the understanding by learners after the speed of auditory lecture increased (Orr et al., 1969; Pastore, 2010; Ritzhaupt, Pastore, & Davis, 2015; Thompson & Silverman, 1977). In fact, in some studies, the level of satisfaction of the participants increased when the auditory speed was increased (Ritzhaupt et al., 2008b). There is not a consensus on the impact of the auditory speed of learning lectures on students' learning. Previous studies did not delve into the impact of time compression on the learning processes.

By far, researchers have examined the relationships between time compressed auditory materials and learning using academic comprehension tests (King & Behnke, 1989; Thompson & Silverman, 1977; Zemlin et al., 1968), cognitive load tests (Pastore, 2012; Pastore, 2010), and satisfaction tests (Ritzhaupt et al., 2008a; Ritzhaupt et al., 2015). However, they used more subjective methods, and collected little objective data such as physiological data. Further, previous studies focused on immediate recalls after listening to auditory learning materials (King & Behnke, 1989) or on training students over time to see if they could adapt to time-compressed speech (Banai & Lavner, 2012; Gabay, Karni, & Banai, 2017). Little research has examined students' delayed memory or recalls of auditory information, which would better reflect the students learning.

In this study, we examined the effects of time-compressed lectures on the individual students' attention, cognitive load, and comprehension of the lectures. In addition to using pre- and post-tests to assess the students' comprehension and the established cognitive load questionnaire to assess the students' perceived loads, we used an electroencephalography (EEG) device to capture the students' brainwave data and understand their attention and meditation (relaxation) values when they were listening to the auditory lectures at different speeds. The following research questions guided the study: (1) Are there any differences in students' attention as detected by the EEG device among the three different auditory speed conditions? (2) Are there any differences in students' cognitive load among the three different auditory speed conditions? and (3) Are there any differences in students' comprehension and memory among the different auditory speed conditions?

2. Literature review and related work

2.1. Auditory learning

Auditory learning has become increasingly popular with the increasing demands for mobile, online, and multimedia learning (Cheon, Lee, Crooks, & Song, 2012; Moreno & Mayer, 2002). The use of auditory playback software, such as podcasts, as a teaching tool has increased dramatically (Pastore, 2010). Students have shown a positive attitude towards integrating these tools into classroom instructions (Evans, 2008).

As for auditory learning, listening comprehension has been a main focus for researchers (King & Behnke, 1989). Comprehending a language being spoken is a complex skill, involving many processes that have become the focus of classroom-oriented research (Call, 1985). Comprehensive listening is typically conceptualized as understanding and remembering a message that is usually associated with long-term memory (Bostrom & Waldhart, 1980). Some researchers defined comprehension as the ability to repeat facts contained in an auditory record, and they developed comprehension tests based on this framework (King & Behnke, 1989).

For comprehension tests, researchers often use both immediate recalls and delayed recalls (Folkard, Monk, Bradbury, & Rosenthall, 1977; Lawson & Hogben, 1998). The immediate recall comprehension tests are usually given immediately after the learners have completed the auditory or reading materials. The delayed recall comprehension tests are usually given some time (usually a week) after the learners have completed the materials. Previous studies showed that the results of immediate recalls and delayed recalls often differed (Folkard et al., 1977), as the learning material would not always remain in short-term memory long enough to be encoded or organized. Yet, as the ultimate goal of education is to pursue long-term memory and learning benefits (Lawson & Hogben, 1998), the delayed recalls should not be ignored in the research of comprehensive tests. Therefore, this study incorporated delayed recall tests.

2.2. Time-compressed instructions or lectures for learning

Research using time-compressed speech dates back to the 1950s (Fairbanks, Guttman, & Miron, 1957). When researchers began experimenting with compressed sounds, they changed the pitch and rhythm. As a result, they changed the sound quality, often making the auditory sound like a fast-paced chipmunk voice, which was distracting to students (Pastore, 2010). As algorithms have been improved, researchers have found ways to artificially shorten the duration of the auditory signal without effecting the fundamental frequency of the signal (Golomb, Peelle, & Wingfield, 2007). The time compression used in this study increased the auditory rate without changing the auditory quality.

Many researchers reported that auditory lectures that were sped up a little bit did not have a significant negative impact on learning. Orr et al. (1969) found that the auditory material with spoken speed accelerated for 1.5x times had no significant difference on the listeners' quiz choices and understanding of the learning material as compared to the material in its original speed. Ritzhaupt et al. (2008b) investigated the effects of different auditory speeds on learners' performance and satisfaction. The authors set the compression speed to 1.0x, 1.4x, and 1.8x of the original multimedia presentation and found that there was no significant difference in performance among the different conditions. However, there were significant differences in the learner' satisfaction levels. The results showed that the learners in the 1.4x condition had the highest satisfaction scores.

Some studies have shown that an increase in compression rate led to an increase in learning difficulties and an adverse effect on learning. Zemlin et al. (1968) had 40 college students assess the difficulty levels of the auditory materials at different compression speeds (1.2x, 1.3x, 1.6x, and 2x). The results showed that starting from 1.2x times, the students' perceived difficulty levels of the materials increased. When the time was compressed at 50% (i.e., doubling the speed of the original spoken materials), the students' perceived difficulty level of the material reached at about 5 times of the difficulty level of the original material. Ritzhaupt and Barron (2008a) found that the scores of learners' content recognition was significantly reduced at very fast speed (2.5 times). Ritzhaupt et al. (2015) found that increased speed (up. To. 1.5x) did not affect learners' comprehension of the listening materials, but the learners' satisfaction declined.

Existing research shows that multimedia auditory lectures can be compressed to a certain extent, and such processing may not cause much loss of information, and sometimes may facilitate the learning of auditory materials. Yet, when the compression ratio rises to a certain level, the compressed auditory will have a significant negative impact on the students (King & Behnke, 1989). There have been studies on time-compressed speech, mainly focusing on the learner's learning performance (Adank & Janse, 2009), satisfaction level (Ritzhaupt & Barron, 2008a), and cognitive load (Pastore, 2012; Pastore, 2010), but there is a lack of research on the effects of time compression on attention, cognitive load, and long-term memory.

2.3. Cognitive load, attention, and auditory learning

In the learning process, if learners do not know the limitations of their working memory or do not adopt complex problem-solving strategies, they may be subject to learning interference and suffer from cognitive overload (Sun & Yu, 2019). There are three types of cognitive load: intrinsic, extraneous, and germane loads (Sun & Yu, 2019; Sweller, 1988). Intrinsic cognitive load is the load placed on working memory from task-inherent complexity of the materials to be learned (Ayres, 2006). Extraneous load refers to the cognitive load caused by the way information is presented and the requirements of teaching activities (Künsting, Wirth, & Paas, 2011). The germane load refers to the mental resources required for acquiring and automating schemata in long-term memory, which contributes to students' learning (Debue & Van De Leemput, 2014).

Cognitive load theory suggests that in complex cognitive tasks, learners who are overwhelmed by a large number of interactive information elements would not be conducive to meaningful learning (Van Gog, Paas, & Sweller, 2010). To this end, cognitive load theory focuses on the concentration and use of cognitive resources in learning and problem solving (Chandler & Sweller, 1991) to keep working memory in the right amount without overloading it. In two studies, Pastore explored the association between time compression techniques and learners' cognitive load levels (Pastore, 2012; Pastore, 2010). One study investigated the impact of a measurement chart and time-compression teaching on learners' perceived cognitive load (Pastore, 2010). The other study measured the impacts of time compression teaching and redundancy (with text) on learning and learners' perceived cognitive load (Pastore, 2012). Both studies showed that the cognitive load of learners did not increase with a little increase of the speed compression (25%). However, the participants had a higher level of cognitive load and lower level of learning performance when the speed compression rate reached at 50%.

Cognitive load theory focuses on the fact that learning materials occupy the learners' working memory (Sun & Yu, 2019). Studies have shown that cognitive load does not increase significantly in time compression that is not increased too much (Pastore, 2012). Yet, so far, insufficient attention has been paid to the impacts of time compression in the multimedia environment. At present, there is very limited research on the brain states of learners when they listen to the time-compressed auditory lectures. Some researchers used functional magnetic resonance imaging (fMRI) techniques to investigate the responses of spoken and written sentences in the brain. The results showed that the activation rate and amplitude of the cortical region were significantly different (Vagharchakian, Dehaene-Lambertz, Pallier, & Dehaene, 2012). This study explores the changes in learners' cognitive load at higher speeds and adds more physiological evidence to explain this important issue.

Brainwave detection technologies such as the Electroencephalography (EEG) are usually used to detect abnormalities in people's brain waves or electrical activities. In research, EEG has been used to study cognitive development and activities such as studies of time-compressed auditory learning. Attention and meditation values are two important indicators of brain wave measurements. Smith, Colunga, and Yoshida (2010) pointed out that learning depends on attention, and attention plays an important role in aggregating, acquiring, and applying knowledge in daily lives. Learners need to stay in a highly concentrated state, but excessive concentration may also have a negative effect. Concentration can benefit from meditation or relaxation, both of which can help people calm down and recharge their attention and energy (Hsu, 2017). NeuroSky, which is a popular mobile brainwave EEG sensor, has been shown to be effective in the measuring attention and meditation values (Lin, Su, Chao, Hsieh, & Tsai, 2016; Liu, Chiang, & Chu, 2013; Sun et al., 2018). This study used NeuroSky to explore the effects of auditory speed on the participant's attention and meditation. Through collecting the brainwave data of the students while listening to the auditory lectures at three different speeds, we attempt to gain a deeper understanding of the effect of time-compressed auditory lectures on learning.

3. Method

This study aims to examine how time compression influences individual students' attention, cognitive load, understanding and memory of the lectures. We used pre- and post-tests and questionnaires to assess students' learning and cognitive load. We used the EEG device to capture the students' brainwave data and understand their attention and meditation/relaxation values while they were listening to the auditory lectures at different speeds. The study explored whether there were any differences in students' attention, cognitive load, comprehension, and memory among the three different auditory speed groups.

3.1. Participants and auditory lectures/materials

The participants included 25 graduate students in China. Fifteen students (60%) of the participants were female, and the average age of the participants was 24.8 years old. A total of three auditory lectures, equivalent in content and difficulty levels, were selected for the experiment. Each lecture included about 2900-3000 written Chinese characters. All lectures were narrated with the same male voice. The article came from three popular books. The pronunciation, intonation, and depth of interpretations were also consistent. The lengths of the three auditory lectures were about 10 minutes at the normal (1.0x) speed. For the study, we modified two of the lectures in the way that they would be at faster speeds, i.e., 1.5x and 3.0x speeds. As a result, we had one auditory lecture, entitled "Mastering the skills of practice" spoken at the normal speed (1.0x); one lecture entitled "Office designs for creativity" spoken at the fast speed (1.5x); and one lecture entitled "Peek performance" spoken at the very fast speed (3.0x). As such, the three different voice speeds turned the auditory lectures into three different auditory durations or lengths. The presentation time lengths and approximate Words per Minute (WPM) for the three lecture conditions are provided in Table 1.

Table 1. Auditory speeds, presentation lengths, and the approximate words per minute (WPMs) of the three auditory lectures

Auditory lecture speed conditions	Presentation lengths (minutes: seconds)	Approximate WPMs (in Mandarin Chinese)
Normal (1.0x times)	10:00	298
Fast (1.5x times)	6:40	442
Very fast (3 times)	3:20	894

3.2. Participants and auditory lectures/materials

Before listening to the auditory lectures, the participants completed a demographic survey and a pre-test measuring their prior knowledge of the lecture content. Each participant came to the researcher's lab and participated in the experiment individually at a time. They were asked to listen to all three lectures with different speeds. While they were listening to the auditory lectures, the participants wore the EEG brainwave equipment, which recorded their attention and meditation values. After completing each auditory lecture, the participant took a quiz (post-test) of the lecture and a cognitive load survey. These processes were repeated for the 2nd and the 3rd lecture and assessments. A week after the experiment, each participant was tested of their knowledge of the lectures again. Figure 1 below shows the experimental procedure.

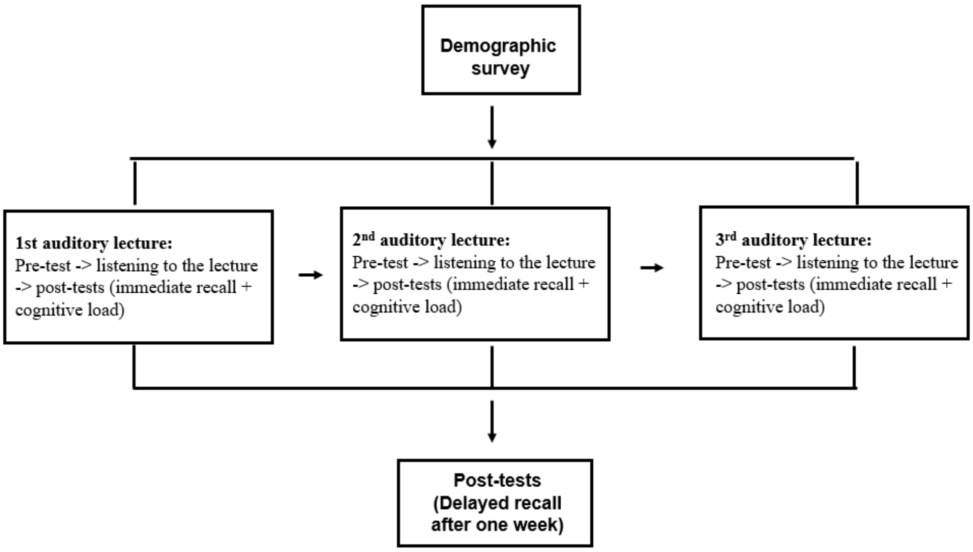


Figure 1. Experimental procedure

The three auditory lectures were purposefully sequenced in the way that when the first participant came in and followed in the sequence of “normal > fast > very fast” speed, the second participant would listen to the lectures in the sequence of “fast > very fast > normal” speeds, and the third participant would listen to the lectures in the sequence of “very fast > normal > fast” speeds. Consequently, there were six different sequences randomly assigned to the students coming to the lab to participate in the study. The total time each participant spent to complete the study (listening to the lectures and completing tests and questionnaires) was about 40 minutes.

3.3. Instruments

3.3.1. EEG brainwave detection system

NeuroSky headset recorded electroencephalogram (EEG) data through a single touch sensor on the forehead of the learner. The eSense is a NeuroSky’s proprietary algorithm for representing mental states. To calculate eSense, the NeuroSky Think Gear technology intensifies the raw brainwave signal and removes the ambient noise and muscle movement. The eSense algorithm is then applied to the remaining signal, resulting in explicated eSense meter values. Based on real-time EEG data, the headset could output two values, namely attention and meditation (i.e., relaxation). Both the attention value and the meditation value were between 0 and 100. Previous research has shown that NeuroSky headsets provide sufficient, effective and reliable data for studies of this nature (Hardy, Drescher, Sarkar, Kellett, & Scanlon, 2011; Chen & Huang, 2014). The analysis results showed that the attention value and meditation value measured by NeuroSky headsets had satisfactory validity and reliability.

3.3.2. Learner cognitive load questionnaire

The cognitive load scale, developed by Paas (1992) and Sweller, Merrienboer, and Paas (1998) was adopted to assess the participants’ cognitive loads while listening to the three auditory lectures at different speeds. The scale consisted of eight items, including five items for “mental load” and three items for “mental effort.” A seven-point Likert scale was used. The Cronbach’s alpha values of the two dimensions were .86 and .85, respectively, demonstrating the high reliability of the scale.

3.3.3. Learning performance tests

Three listening comprehension tests were designed based on the three auditory lectures to assess the students’ learning performance. As mentioned earlier, the three lectures were “Mastering the skills of practice” spoken at the normal (1.0x) speed; “Office designs for creativity” spoken at the fast (1.5x) speed; and “Peek performance” spoken at the very fast (3.0x) speed. Ten multiple-choice questions were developed for each auditory lecture, based on Bloom’s Taxonomy (Bloom, 1956). Five out of 10 questions were basic-level questions (related to knowledge recall and understanding); three out of 10 were intermediate-level questions (related to knowledge application and analysis); and two were in-depth level questions (related to knowledge evaluation and creation).

The highest score a participant could receive was 10 points (with 1 point for each correct answer). Basic level questions sought information or facts that would be easily recalled from the auditory lecture. Intermediate-level comprehension questions required the participants to summarize the key concepts (or to discard untrue statements). In-depth questions required the participants to assess and synthesize what's said and create new meanings. Responses to these questions allowed us to see the depth of comprehension by the participants, and to examine the impact of the auditory lectures at different speeds on the participants.

The participants received one point for each question answered correctly in the quiz, with a total score of 10 for one lecture. Based on the statistical analysis of a pilot study, we confirmed that the difficulty levels between the three tests were comparable. The three tests had high reliability in assessing academic performance. The three tests were each used three times during the experiment. First, each test was used as a pre-test to assess the participants' prior knowledge of the lecture to be listened. Then the test was used immediately after the participant listened to the lecture to assess immediate recall of the lecture. Finally, after a week, the tests were used to assess the participants delayed recalls of the lectures. In order to ensure the reliability and effectiveness of the return visit after one week, the participants was not informed at the beginning that they would be tested again a week later.

4. Results

4.1. Analysis of individual attention and meditation from EEG data

ANOVA was used to answer the first research question, "Are there any differences in students' attention as detected by the EEG device among the three different auditory speed conditions?" The results of the individual attention and meditation of the three conditions are shown in Table 2.

Table 2. ANOVA results of attention for the three auditory lectures

Auditory lecture speed conditions	*N*	Mean	*SD*	*F*	Post hoc tests
Normal (1.0x times) (a)	25	46.39	8.63	6.263^{**}	(c) > (a)
Fast (1.5x times) (b)	25	45.32	6.00		(c) > (b)
Very fast (3 times) (c)	25	52.76	9.13		

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

According to the ANOVA result for attention (F = 6.263, p < .001), the students had a significant higher attention in the very fast speed condition than when they were in the normal speed (p < .05) and the fast speed (p < .05) conditions. There was no significant difference in the attention value between the normal speed and the fast speed conditions (p > .05). When it comes to meditation, the students in the very fast speed condition had a significantly lower meditation than when they were in the other two conditions (p < .05) (see Table 3).

Table 3. ANOVA results of meditation (relaxation) of the three auditory lectures

Auditory lecture speed conditions	*N*	Mean	*SD*	*F*	Post hoc tests
Normal (1.0x times) (a)	25	55.09	9.26	4.383^{*}	(c) < (a)
Fast (1.5x times) (b)	25	54.07	6.41		(c) < (b)
Very fast (3 times) (c)	25	48.80	8.20		

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

4.2. Analysis of individual cognitive load

ANOVA was also used to answer the second research question, "Are there any differences in students' cognitive load among the three different auditory speed conditions?" The ANOVA results regarding the cognitive load of the three conditions are shown in Table 4.

There were significant differences in cognitive loads between the three conditions (F = 35.11; p < .001). As shown in Table 4, the students reported a significantly higher cognitive load in the very fast speed condition than in the normal speed condition (p < .05) and the fast speed condition (p < .05).

Table 4. ANOVA results of the reported cognitive loads of the three auditory lectures

Auditory lecture speed conditions	*N*	Mean	*SD*	*F*	Post hoc tests
Normal (1.0x times) (a)	25	2.27	1.54	35.11^{***}	(c) > (a)
Fast (1.5x times) (b)	25	2.34	1.22		(c) > (b)
Very fast (3 times) (c)	25	5.24	1.50		

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

4.3. Analysis of the learning performance

By using the paired-sample *t* test, we examined whether the three lectures generated good learning performance based on pretest and posttest scores. In addition, we examined whether the learning performance differed significantly among the three auditory conditions. The analytical results indicated that the students did significantly better in their immediately recall post-tests than in the pre-tests in all the three lectures (see Table 5).

Table 5. Paired-sample *t* test results of pre-tests and immediate recall post-tests with three different time-compressed lectures

Auditory lecture speed conditions	Learning of the lectures	*N*	Mean	*SD*	*t*
Normal (1.0x times) (a)	Pretest	25	3.36	1.08	16.65^{***}
	Posttest	25	9.40	1.15	
Fast (1.5x times) (b)	Pretest	25	3.68	1.18	19.77^{***}
	Posttest	25	9.00	1.00	
Very fast (3 times) (c)	Pretest	25	3.80	1.22	8.414^{***}
	Posttest	25	6.56	1.26	

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

Before analyzing the learning performance of the three conditions, we did a baseline analysis of the participants' the pretest scores to answer the third research question, "Are there any differences in students' comprehension and memory among the different auditory speed conditions?" The ANCOVA results regarding the learning performance in the three conditions are shown in Table 6.

Table 6. ANCOVA results of immediate recall post-test comparisons between the three auditory lectures

Group		Pre-test		Post-test		ANCOVA	
	N	Mean	*SD*	Mean	*SD*	*F*	Pairwise comparison
Normal (1.0x times) (a)	25	3.36	1.08	9.41	1.15	38.9254^{***}	(a) > (c)
Fast (1.5x times) (b)	25	3.68	1.18	9.00	1.00		(b) > (c)
Very fast (3x times) (c)	25	3.80	1.22	6.55	1.26		

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

According to the ANCOVA results of learning performance ($F = 38.9254$, $p < .001$), the average scores of the immediate recall post tests were 9.41, 9.00, and 6.55 for the conditions at the normal speed (1.0x), with fast speed (1.5x), and with very fast speed (3.0x), respectively. Students in the very fast speed condition (3x times) group had significantly lower listening comprehension scores (6.55) after the post-hoc test ($p < .001$) than the fast group (9.00) and the normal group (9.41). There were no differences between the fast and the normal groups ($p > .05$).

Table 7. ANCOVA results of the delayed recall (after one-week) post-test comparisons between the auditory lectures

		Pretest		Delayed recall		ANCOVA	
Group	*N*	Mean	*SD*	Mean	*SD*	*F*	Pairwise comparison
Normal (1.0x times) (a)	25	3.36	1.08	8.09	1.36	19.50^{***}	(a)>(c)
Fast (1.5x times) (b)	25	3.68	1.18	8.02	1.54		(b)>(c)
Very fast (3x times) (c)	25	3.80	1.22	5.45	1.87		

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

Table 7 shows the ANCOVA results of learning performance (after one-week) ($F = 19.50$, $p < .001$). The mean values of the delayed-recall (after one week) post-test scores were 8.09 for the normal speed, 8.02 for the fast speed, and 5.45 for the very fast speed. The students in the very fast speed condition (3x times) had significantly lower listening comprehension scores (5.45) after one week ($p < .001$) than the fast group (8.02) and the normal group (8.09). There was no significant difference between the fast speed and the normal speed conditions ($p >$ .05).

5. Discussions, limitations, and conclusion

The study has several significant implications for educational theory and practices. First, we found that there was no significant difference in students' attention, cognitive load, comprehension, and memory between the normal speed (1.0x) and the fast speed (1.5x) conditions. The findings on learning performance are consistent with previous studies that examined the impact of time compression on learning outcomes (Adank & Janse, 2009; Thompson & Silverman, 1977). In addition to immediate recalls, we added delayed recalls of the lectures. The results showed no significant difference in the students' delayed recall scores between the normal (1.0x) speed and the fast (1.5x) speed. This dimension added weight to the claim that increasing speed of the auditory lectures to a degree may not affect their learning negatively. Further, we used the EEG brainwave device and captured the students' attention during their learning processes, and we asked the students to report their cognitive load for each lecture they listened. The results showed that the students exhibited similar levels of attention, meditation / relaxation, and cognitive load in the normal (1.0x) speed and the fast (1.5x) speed conditions. Based on all these findings, we can safely suggest that increasing the auditory lecture speed up to 1.5x times might not negatively affect students' learning, attention or cognitive load. This is an important finding because it shows that learners can increase their learning efficiency by speeding up their auditory lecture to a certain degree. At the same time, educators can potentially adopt this strategy in teaching.

Second, the results of this study showed that when the speed was increased three times (3.0x), the students' learning performance suffered greatly. Further, the lecture at the 3.0x speed significantly increased the students' cognitive load. These are consistent with prior studies as well (Ayres, 2006; Künsting et al., 2011). The participants' brainwaves detected by EEG device also showed that the participants' attention was three times more intensive, and that their meditation values were significantly lower than when they were in the other two conditions (the normal and the fast speed conditions). That is, when the participants were listening to the auditory lecture at the very fast (3.0x) speed, they were very intense and not-relaxed, although they were at a very high level of attention. This research finding can be used to further explain the combined effects of the fast (3.0x) speed on the learning process. In a more stressful state, even if the attention is more concentrated, the cognitive load is more likely to increase, and the learning outcome is worse. This finding provides more evidence for understanding the speeds to which to compress the instructional videos without sacrificing the learning (Ritzhaupt et al., 2015). Therefore, instructional designers and learners should not simply pursue faster speed and obtaining information in a shorter period of time, but instead, they should choose an appropriate time compression ratio, to ensure that meaningful learning can take place (Mayer, 2003).

The study has some limitations. First, although we assessed the participants' prior knowledge of the auditory lectures, we did not focus on the individual differences pertaining to attention, cognitive load, past experiences of time-compression lectures, or memory capacities. Future studies could look more in-depth into individual differences. In addition, the use of the EEG device limited the amount of time for experiments. The total time each participant spent to complete the study was about 40 minutes. The short lectures and the extended time to complete the study could all affect the results of the study. Last but not the least, although we purposely sequenced the three different speed lectures in the way that they would take turns to be the 1st, 2nd, or the 3rd lectures to be listened and completed by different participants, the sequence might still have affected the participants' attention, cognitive load, comprehension, and memory.

Despite these limitations, this study addressed several research gaps. We collected physiological data of brain waves to better understand the students' learning processes of auditory lectures at different time-compression speeds (Banai & Lavner, 2012; Gabay et al., 2017; Pastore, 2012). Compared to prior studies, this study used EEG data and added the delayed recall assessments to examine to what extent the students were able to retain the information after one week. The results of the study help researchers, educators, and learners further understand the effects and underlying mechanisms of time-compressed auditory learning. This study further confirms that a certain degree of time compression may be acceptable and may not affect learners' attention and meditation values, cognitive load, or learning outcomes negatively. As digitally recorded auditory learning such as podcasts is widely integrated into multimedia learning environments (Evans, 2008; Moreno & Mayer, 2002), educators and learners can choose appropriate time compression ratios to increase learning efficiency (Littlejohn, Hood, Milligan, & Mustain, 2016). Once the time compression ratio is found to be too high and the learner perceives tension, educators can reduce the time compression ratio. The widespread use of time compression for auditory lectures highlights the value of this research. Time compression is becoming a new habit for more and more learners to obtain information on multimedia, mobile and online learning environments (Pastore, 2012; Pastore, 2010). Major media and learning platforms can be optimized in terms of time compression ratio settings, providing learners with better time compression options (Gillani & Eynon, 2014). Future researchers should continue to explore areas where the theory and practices are closely integrated.

References

Adank, P., & Janse, E. (2009). Perceptual learning of time-compressed and natural fast speech. *The Journal of the Acoustical Society of America, 126*(5), 2649-2659.

Ayres, P. (2006). Impact of reducing intrinsic cognitive load on learning in a mathematical domain. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 20*(3), 287-298.

Banai, K., & Lavner, Y. (2012). Perceptual learning of time-compressed speech: More than rapid adaptation. *PloS one, 7*(10), e47099.

Barabasz, A. F. (1968). A Study of recall and retention of accelerated lecture presentation. *Journal of Communication, 18*(3), 283-287.

Barron, A. E., & Kysilka, M. L. (1993). The Effectiveness of digital audio in computer-based training. *Journal of Research on Computing in Education, 25*(3), 277-289.

Bloom, B. S. (1956). *Taxonomy of educational objectives. Vol. 1: Cognitive domain*. New York, NY: McKay.

Bostrom, R. N., & Waldhart, E. S. (1980). Components in listening behavior: The Eole of short-term memory. Human Communication Research, 6(3), 221-227.

Broadbent, J. (2017). Comparing online and blended learner's self-regulated learning strategies and academic performance. *The Internet and Higher Education, 33*, 24-32.

Call, M. E. (1985). Auditory short-term memory, listening comprehension, and the input hypothesis. *Tesol Quarterly, 19*(4), 765-781.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and instruction, 8*(4), 293-332.

Chen, C. M., & Huang, S. H. (2014). Web-based reading annotation system with an attention-based self-regulated learning mechanism for promoting reading performance. *British Journal of Educational Technology, 45*(5), 959-980.

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

Debue, N., & Van De Leemput, C. (2014). What does germane load mean? An Empirical contribution to the cognitive load theory. *Frontiers in psychology, 5*, 1099-1111.

Evans, C. (2008). The effectiveness of m-learning in the form of podcast revision lectures in higher education. *Computers & Education, 50*(2), 491-498.

Fairbanks, G., Guttman, N., & Miron, M. S. (1957). Auditory comprehension of repeated high-speed messages. *Journal of Speech and Hearing Disorders, 22*(1), 20-22.

Folkard, S., Monk, T. H., Bradbury, R., & Rosenthall, J. (1977). Time of day effects in school children's immediate and delayed recall of meaningful material. *British Journal of Psychology, 68*(1), 45-50.

Gabay, Y., Karni, A., & Banai, K. (2017). The Perceptual learning of time-compressed speech: A Comparison of training protocols with different levels of difficulty. *PloS one, 12*(5), e0176488.

Gillani, N., & Eynon, R. (2014). Communication patterns in massively open online courses. *The Internet and Higher Education, 23*, 18-26.

Goldhaber, G. M. (1970). Listener comprehension of compressed speech as a function of the academic grade level of the subjects. *Journal of Communication, 20*(2), 167-173.

Golomb, J. D., Peelle, J. E., & Wingfield, A. (2007). Effects of stimulus variability and adult aging on adaptation to time-compressed speech. *The Journal of the Acoustical Society of America, 121*(3), 1701-1708.

Hardy, J. L., Drescher, D., Sarkar, K., Kellett, G., & Scanlon, M. (2011). Enhancing visual attention and working memory with a web-based cognitive training program. *Mensa Research Journal, 42*(2), 13-20.

Hsu, W. Y. (2017). A Wireless brainwave-driven system for daily-life analyses and applications. *Telematics and Informatics, 34*(8), 1793-1801.

King, P. E., & Behnke, R. R. (1989). The Effect of time-compressed speech on comprehensive, interpretive, and short-term listening. *Human Communication Research, 15*(3), 428-443.

Koroghlanian, C. M., & Sullivan, H. J. (2000). Audio and text density in computer-based instruction. *Journal of Educational Computing Research, 22*(2), 217-230.

Kress, G., & Selander, S. (2012). Multimodal design, learning and cultures of recognition. *The Internet and Higher Education, 15*(4), 265-268.

Künsting, J., Wirth, J., & Paas, F. (2011). The Goal specificity effect on strategy use and instructional efficiency during computer-based scientific discovery learning. *Computers & Education, 56*(3), 668-679.

Lawson, M., & Hogben, D. (1998). Learning and recall of foreign-language vocabulary: Effects of a keyword strategy for immediate and delayed recall. *Learning and Instruction, 8*(2), 179-194.

Lee, P. S., & Chan, A. H. (2003). Chinese speaking times. *International Journal of Industrial Ergonomics, 31*(5), 313-321.

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.

Littlejohn, A., Hood, N., Milligan, C., & Mustain, P. (2016). Learning in MOOCs: Motivations and self-regulated learning in MOOCs. *The Internet and Higher Education, 29*, 40-48.

Liu, N. H., Chiang, C. Y., & Chu, H. C. (2013). Recognizing the degree of human attention using EEG signals from mobile sensors. *Sensors, 13*(8), 10273-10286.

Mayer, R. E. (2003). Elements of a science of e-learning. *Journal of Educational Computing Research, 29*(3), 297-313.

Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology, 94*(1), 156-163.

Orr, D. B., Friedman, H. L., & Graae, C. N. (1969). Self-pacing behavior in the use of time-compressed speech. *Journal of Educational Psychology, 60*(1), 28-31.

Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A Cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429-434.

Pastore, R. (2012). The Effects of time-compressed instruction and redundancy on learning and learners' perceptions of cognitive load. *Computers & Education, 58*(1), 641-651.

Pastore, R. S. (2010). The Effects of diagrams and time-compressed instruction on learning and learners' perceptions of cognitive load. *Educational Technology Research and Development, 58*(5), 485-505.

Ritzhaupt, A. D., & Barron, A. (2008a). Effects of time-compressed narration and representational adjunct images on cued-recall, content recognition, and learner satisfaction. *Journal of Educational Computing Research, 39*(2), 161-184.

Ritzhaupt, A. D., Gomes, N. D., & Barron, A. E. (2008b). The Effects of time-compressed audio and verbal redundancy on learner performance and satisfaction. *Computers in Human Behavior, 24*(5), 2434-2445.

Ritzhaupt, A. D., Pastore, R., & Davis, R. (2015). Effects of captions and time-compressed video on learner performance and satisfaction. *Computers in Human Behavior, 45*, 222-227.

Smith, L. B., Colunga, E., & Yoshida, H. (2010). Knowledge as process: contextually cued attention and early word learning. *Cognitive Science, 34*(7), 1287-1314.

Sun, J. C. Y., Hwang, G.-J., Lin, Y. Y., Yu, S. J., Pan, L. C., & Chen, A. Y. Z. (2018). A Votable concept mapping approach to promoting students' attentional behavior: An Analysis of sequential behavioral patterns and brainwave data. *Educational Technology & Society, 21*(2), 177-191.

Sun, J. C. Y., & Yu, S. J. (2019). Personalized wearable guides or audio guides: An evaluation of personalized museum guides for improving learning achievement and cognitive load. *International Journal of Human–Computer Interaction, 35*(4-5), 404-414.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science, 12*(2), 257-285.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296.

Thompson, A. J., & Silverman, E.M. (1977). Children's comprehension of time-compressed speech: Effect of speaker's familiarity. *Perceptual and Motor Skills, 45*(3_suppl), 1253-1254.

Vagharchakian, L., Dehaene-Lambertz, G., Pallier, C., & Dehaene, S. (2012). A Temporal bottleneck in the language comprehension network. *Journal of Neuroscience, 32*(26), 9089-9102.

Van Gog, T., Paas, F., & Sweller, J. (2010). Cognitive load theory: Advances in research on worked examples, animations, and cognitive load measurement. *Educational Psychology Review, 22*(4), 375-378.

Wang, Q., Ren, X., Hu, J., Li, Q., Cui, S., & Zou, Y. (2018). Preliminary study on reading speed test with IReST for normally-sighted young Chinese readers. [Zhonghua yan ke za zhi] *Chinese Journal of Ophthalmology, 54*(2), 120-124.

Wang, Y. S., Wu, M. C., & Wang, H. Y. (2009). Investigating the determinants and age and gender differences in the acceptance of mobile learning. *British Journal of Educational Technology, 40*(1), 92-118.

Zemlin, W. R., Daniloff, R. G., & Shriner, T. H. (1968). The Difficulty of listening to time-compressed speech. *Journal of Speech and Hearing Research, 11*(4), 875-881.

Lo, Y.-H. G., Lee, S.-Y., & Tien, W.-C. F. (2020). Transforming a Magazine into a Video Involving a Target Audience: A Multiliteracies Case Study in an EFL Context. *Educational Technology & Society*, *23* (3), 27–39.

Transforming a Magazine into a Video Involving a Target Audience: A Multiliteracies Case Study in an EFL Context

Yi-Hsuan Gloria Lo, Sy-Ying Lee* and Wan-Chu Finny Tien
National Taiwan University of Science and Technology, Taiwan // glorialo@mail.ntust.edu.tw // syying.lee@mail.ntust.edu.tw // B10517007@mail.ntust.edu.tw
*Corresponding author

(Submitted February 25, 2020; Revised April 28, 2020; Accepted July 17, 2020)

ABSTRACT: In this article, we presented a case study in an EFL context that investigated how a magazine was transformed into a digital video involving a target audience. A group of fifty international students responded to a survey questionnaire developed based on the multimodality framework and some were interviewed to express their preferences for the content and format of the video and to evaluate different versions of the video product. The results show that the transforming process consisted of four stages including (1) collecting target audience's preferences for the video content and format, (2) converting the discourse type from textual to oral, (3) creating multimodal materials for the video, and (4) (re-)composing the video. The target audience's responses revealed that effective multimodal orchestration could provide a better engagement and viewing experience for the target audience. The multiliteracies competency of the video creators and viewers was deepened and expanded through the digital transforming processes and interdisciplinary collaboration, which enabled EFL learners to experience, conceptualize, analyze, and apply the learned and new knowledge. With the ultimate goal to cultivate EFL learners to become multimodal literate citizens in the global society, this study advances our understanding of multimodality and yields significant pedagogical implications for multiliteracies education and educational technology in the EFL context.

Keywords: Video, Target audience, Multimodality, Multiliteracies, Educational technology

1. Introduction

Literacy has long been defined as the ability to communicate and make sense of the world effectively through language. However, the evolution of technology has created a multimodal world, inundated with websites (e.g., YouTube, Google) and social media (e.g., Facebook, Instagram, Twitter), thus leading us into the realm of new media literacy (Hung, 2019; Lin, Li, Deng, & Lee, 2013). Being literate is no longer solely about being able to read and write, the linguistic mode that has been the primary emphasis in most EFL contexts. A literate being in this 21st multimodal world should be equipped with the ability to "understand the effect of all modes of communication that are co-present in any text" (Kress, 2000, p. 337) including visual, aural, spatial and gestural modes, in addition to linguistic mode (New London Group, 1996). This notion suggests the importance of multiliteracies, referred to as the abilities and competencies "to create, critique, analyze and evaluate multimedia texts" (NCTE, 2013). In response to the trend of multimodal literacies, many composition scholars, such as Alexander and Rhodes (2014), believe that "engaging multimodality is a pressingly necessary task for a wide variety of composition" (p. 71). To advance our understanding of the multimodal composing process, the case study echoes this pressing need by examining how meaning-making in a paper format, the most common form of reading and writing production in EFL contexts, can be transformed into a video via digital multimodal composing (DMC) (Jiang, 2017), a semiotic process using digital tools to create multimodal texts combining different modes. With the goal to communicate with a specified target audience via the video to reach out to a larger population (Mills, 2015), this case study aims to cultivate EFL learners to become multimodal literate citizens in the 21st global society.

How can multiple modes be best "orchestrated" or "combined?" This issue may be informed and addressed by the two important questions raised by the 21st century literacy framework (NCTE, 2013). First, "Do students publish in ways that meet the needs of a particular, ***authentic audience***?" Second, "Do students solve real problems and share results with ***real audience***?" The role of the target audience has been explored in writing composition to some extent (Kakh, Mansor, & Zakaria, 2014; Wong, 2005). However, most of the target audiences in previous research (e.g., Cho & Choi, 2018; Gunel, Hand, & McDermott, 2009; Martinez-Insua, 2019) were not "real" or "authentic;" rather, they were hypothetical because the writers were instructed to imagine a group of audience they were writing to, who may or may not exist. Furthermore, the instructors in most research were often the evaluators to assess and evaluate the final production rather than the target audience. The role of a specified target audience in the digital multimodal process is far scanter and worth investigation.

In this case study, we had two goals: first, to document the process of transforming a magazine (from static images) into a multimodal video (to moving images) for a specific target audience and second, to collect our target audience's responses to refine the "orchestration" of multiple modes. Specifically, our goals led us to ask two research questions: (1) What is the evolving process of transforming a magazine into a multimodal video when a target audience is involved? (2) What can our target audience's responses to the videos inform us about the orchestration of the multiple modes when transforming a magazine into a video?

2. Literature review

2.1. The orchestration process involved in multimodal literacies

In multimodal literacies, language is acknowledged as "only one of several semiotic tools for communicative purposes" (Yeh, 2018, p. 29). While digital technology has been identified as a powerful mediation to expand and enhance multimodal meaning-making among people nowadays, including language learners, pedagogies integrating digital and conventional literacies have been called for (Hafner, 2014; Jewitt, 2006; Kress, 2003; Mills, 2010). In response to this, many studies have investigated digital multimodal composing (DMC) by examining how students produce digital videos (Campbell & Cox, 2018; Hafner, 2014; Jiang, 2018; Yeh, 2018). The three following studies were selected for further discussion because of the close relevancy they had with our current case study. Each of them reported on how students created videos by following certain steps or processes when producing the video.

Firstly, Hafner (2014) discovered six steps in video making: reading, data collection, scripting and storyboarding, filming, editing, and sharing. Following these steps, his students identified "a wide range of semiotic resources, including moving images and animation, charts and tables for scientific data, subtitles, different camera angles and lighting, background music, sound effects, interesting locations, interesting participants, and facial expression" (p. 669). Results showed that one main source of rhetorical challenge for his students during video production is multimodal orchestration. For instance, the students were concerned if "over-relying on multimedia" or "excess use of technique may annoy the audience." Another challenge was "how to write an interesting script for narration for the documentary" because they viewed "language as an equally important resource" (p. 668). However, the students did not assess the effectiveness of the visuals and script, which is believed to be an important procedure missing in the orchestration process.

In a similar vein, Yeh (2018) mapped out three steps to create a video among EFL college students: composing the scripts, enacting the scripts, and editing the videos. In this study, the students first read information for their selected topics and composed their scripts for their videos. To enact their scripts, the students had to "come up with different innovative ways of presenting their topics through the combination of multiple modes to present their core themes" (p. 30). In the editing stage, they managed to "combine multiple modes such as adding text, pictures, subtitles, effects, narrations, soundtracks, and PowerPoint slides, to tie all their ideas together to construct the videos" (p. 31). While several perceived benefits were found in language acquisition (e.g., vocabulary, speaking, translation and writing), cultural learning, and multimodal capacities (e.g., editing skills for multimodality), how multiple modes were orchestrated or combined to create a multimodal video, however, was not explored in detail.

Different from Hafner's and Yeh's studies which mainly focused on the production processes students underwent, Monte Mŏr (2015) detailed what and how college students in Brazil experienced and learned through video creation using the *"Learning by Design"* framework proposed by Kalantzis and Cope (2005) to theorize how students learn when they engage in digital media production. This framework outlines four knowledge processes to support teaching and learning: experiencing the known and the new, conceptualizing by naming and theorizing, analyzing functionally and critically, and applying appropriately and creatively. Students in Monte Mor's study recorded and compiled their discussions on the educational system and entrance examination they had all experienced. Students performed their ideas using their own voice and facial expressions for real viewers, their peers. They critiqued and evaluated the system based on *the known* and learned *the new* by viewing and discussing others' experiences and arguments. This project was made authentic, unique, and original because of the application of the technical tool, the *Flash* software, which helped design the interactive display of the multiple videos of all student narrators showing on the screen. They eventually created a collective narrative on education, society, and the future job market as a class project. By the same token, our study intended to use this framework to examine whether these knowledge processes were also experienced by our EFL learners during video production.

2.2. The role of target audience in (video) composing

Research has shown that target audiences play different roles in the composing process and final outcome, resulting in different impacts on the writers themselves as well as the writing outcomes. While the role of target audience is evident in writing composition (the linguistic mode), what would be the role that target audiences can bring in a video composing process and final outcome?

Wong's (2005) study showed that the intended target audiences, although imaginary, gave different mental representations, and thus a wide range of writing strategies were employed to serve different rhetorical purposes at different stages of the composing process for their "imagined" or "hypothetical" target audiences. On the other hand, writing-to-learn for different target audiences can have an impact on the learning of the writers themselves. In investigating writers' conceptual understanding of biology (Gunel, Hand, & McDermott, 2009), it is found that students writing (in a hypothetical context) for peers or younger students performed significantly better on conceptual understanding than students writing for the teacher or parents. Furthermore, in Cho and Choi's (2018) study that investigated the effect of audience specifications on the summary writing, it showed that writers with a specified audience outperformed those without one.

Target audience also plays a role to determine what and how language is represented and organized. Martinez-Insua's (2019) research indicated that texts that address learned audiences tend to be contentful (carrying more weight on content), which demands certain background knowledge from the reader in order to create or infer new discourse, whereas texts presented to lay audiences were more contentlight (less weight on content). It is also found that whereas most subject themes are contentful in formal written texts, most of them are contentlight in informal spoken texts. The distinction in different modes of presentation catering to different target audiences can also be found in Hafner's (2014) students' digital video projects.

In Hafner's (2014) study, students were required to create a multimodal scientific documentary to share with "a general audience of nonspecialists" through YouTube, and a written lab report for "a specialist audience." Although Hafner pointed out that his students met the challenge of writing for "an authentic audience" when "combining a range of modes," it is not clear who the target audience was, what language choices were made for the two types of audiences, and whether or not ways of combing a range of different modes to "appeal to their audience" (p. 655) would result from their specified target audience? The same questions could also be asked in response to Yeh's (2018) project aiming at creating a video for "authentic audiences in online communities" (p. 29). In summary, throughout the video composing process, a lack of specification of the target audience may create difficulty in determining appropriate contents, ways of combining modes appealing to the audience's needs, and assessing whether or not it communicates its intended messages effectively.

3. Methodology

3.1. The context of the case study

A case study approach was particularly useful for exploring the particularity and complexity involved in the processes of transforming a magazine to a video because it "allows in-depth, multi-faceted explorations of complex issues in [the] real-life settings" (Crowe, Cresswell, Robertson, Huby, Avery, & Sheikh, 2011, p. 1), which may not be feasibly possible via other approaches. This case study took place in a one-year capstone course proffered for university English majors to conduct a special project on a self-selected topic as a requirement for graduation. Capstone courses have been practiced worldwide to provide students with a culminating, integrative experience of learning to demonstrate what they have acquired in previous years by synthetically applying the learned skills in simulated or real-world situations (Wagennar, 1993). In this course, with the first two authors also the instructors, students in groups chose to design projects in any form that was deemed appropriate and creative to solve a predetermined problem. Among the eleven projects this year, one video-making project, conducted by the third author and her group members, which this research study was based on, fitted our research interest in multiliteracies as an EFL pedagogy.

3.2. Participants

One of the reasons for the increasing number of international students in Taiwan is that they want to experience Taiwan's culture (Pan & Zhang, 2019). Therefore, international students became the target audience of the study. Prior to the study, an invitation message informing of the several stages of cooperation required was posted on the Internet accessible to the international student body. Among the many who replied and expressed their interest and voluntariness to participate in this research, fifty students from different countries studying in graduate level programs were selected based on three criteria: (1) who stayed in Taiwan for at least a semester; (2) who had expressed a strong interest in Hualien; and (3) who were sufficiently competent in English to understand the survey questions and were able to respond to the video with opinions and suggestions.

3.3. Instruments and data collection

Both qualitative and quantitative data were collected using the following instruments.

Magazine. The magazine sought to offer international tourists an in-depth introduction to Hualien. This magazine, a multimodal project prior to the capstone course, featured three modes: linguistic (carefully-crafted words in print), visual (pictures), and spatial (professional layout and aesthetic cover design).

A survey questionnaire. The survey was designed following the five modes in order to understand the target audience's preferences for the what (content) and the how (format) of the video to be created and was distributed to the 50 international students online (see Appendix A at https://drive.google.com/file/d/1HlAPnZwcCupHX4MBG7Il8NlyU61_VARd/view). The first part contains three sections (e.g., natural landscape, the Japanese colonization period, and aboriginal groups). The participants were asked to choose three options in the first section, two options in the second, and one option in the third according to their personal preferences. The second part investigates the format based on the five modes: linguistic (genre, wording of the subtitles), visual (hosting style, subtitles and special effects), visual and spatial (style), aural (background music, sound effects), gestural (gestures), and others (length).

Interviews. Face-to-face interviews were conducted in English with 10 of the 50 respondents to elicit further elaboration on the questionnaire items. After the two videos were produced, one-on-one interviews were held in English with another 10 international students to further understand the target audience's views on the two versions of the video. The first part of the interview investigated how much (on a rating scale of 1 to 10) an interviewee liked the six tourist spots, and the second part seeing how much (on a rating scale of 1 to 10) an interviewee enjoyed the video presentation (overall style, length, hosting style, subtitles, background music, sound effects, and special effects). Lastly, the interviewees were asked to elaborate on their ratings and the reasons for the differences in the ratings (if any). For example, "How did you rate the two versions of "Qingshui Cliff?" Why did you give the second version a higher rating (Score = 9) than that (Score = 5) in the first version? What were your reasons for the different ratings? (See Appendix B at https://drive.google.com/file/d/1K49MbBKv5wkYigLjzGCkO1JEgqm8upmy/view).

Research trip to Hualien. A three-day research trip was conducted in Hualien to explore the place and culture in person and to shoot the video according to the results of the first survey and interview.

Two versions of the video. Two versions of our video were produced, one by the research team and the other with the assistance of a video-editing professional to elicit the target audience feedback (see the Results).

Field notes. The research team kept field notes to document in detail the process of conducting the case study.

3.4. Data analysis

Descriptive statistics (percentages and means) were firstly obtained from the questionnaires. Next, the interview data, transcribed verbatim and coded using the multimodal framework proposed by the New London Group (1996), were then categorized into themes. Lastly, the qualitative results triangulated with the descriptive statistics were used to address the research questions. Specifically, to examine the transformation process from a magazine to a video, the research team first analyzed the field notes by charting the major stages based on the chronological order, which resulted in the emergence of the four stages (see Results: 4.1). Further data analysis for each stage (except stage II) employed the five modes as an analytical framework to analyze the target audience's preferences (Stage I), the orchestration of the five modes (Stage III), and the differences between the two versions of the video (Stage IV). On the other hand, data analysis for stage II centered on comparing and contrasting the differences between the linguistic texts in the magazine and in the video scripts in relation to formality (Einhorn, 1978), depth (Schallert, Kleiman, & Rubin, 1977), personal references (Strauss, Feiz, & Xiang, 2018), word choice (DeVito, 1965; Miller, 2011), and syntax (Wilkinson, 1971). To address research question 2, when analyzing the target audience's responses to the two versions of the video, the research team first coded qualitative interview data provided by the target audience to analyze the reasons for their preferences, which were then triangulated with the rating results. The results of the analysis were then corroborated with relevant data analysis gathered from the previous stages, based on which three major themes were generated (see Results: 4.2).

4. Findings

4.1. RQ1: What is the evolving process of transforming a magazine into a multimodal video when a target audience is involved?

Grounded in the five modes of multimodality, a four-stage framework for transforming a magazine into a multimodal video emerged, as presented below.

4.1.1. Stage I: Collecting target audience's preferences for the video content and format

In the first part, our target audience expressed *what* they wanted to see in the video. For "natural landscape," the ranking results showed that Qingshui Cliff (40%), Taroko Gorge (38%), and Yanzikou Trail (42%) were ranked first to third, respectively. For "the Japanese colonization period," Pine Garden was ranked first (54%) followed by Ji-An Shrine (40%). For "aboriginal groups," the Amis was ranked first (64%).

In the second part, the target audience were asked about nine aspects related to *how* they wanted the video to be presented. For what type of genre to introduce Hualien, more than half our target audience (52%) hoped a micro film in 7 to 9 minutes long (64%) and supplemented with interesting and humorous elements to be made, so it could be "less boring" (interviewee 5). Most of the target audience (80%) wanted to have a host and expected the host to be a Taiwanese who speaks English and could interact "naturally with the local people" (Interviewee 8). As for whether the subtitles were needed, 62% of the target audience chose "subtitles throughout the video." The subtitle featuring both Chinese and English was more preferred because it looked "more professional" (Interviewees 4, 6, 7, 9). Moreover, the subtitle wording should be simple, fact-based (74%), instead of rich, detail-oriented descriptions. Among the five options of background music, country music was the most popular (60%), because of its "more relaxing tone" (Interviewee 2) and "a brisk rhythmic pace" (Interviewee 8). Interestingly, most of our target audience did not appreciate any special effects (66%) nor sound effects (80%). Collecting the target audience's feedback before producing the video was pioneering as it put the target audience's exact needs into consideration, leading the following orchestration process to be more purposeful, yet more complicated.

4.1.2. Stage II: Converting the discourse type from textual to oral

Based on our participants' preferences, the scripts were composed and featured simple, fact-based oral language. To communicate with a specified target audience in an authentic situation, we looked into the differences between written and oral language as laid out by applied linguists with regards to formality, depth, personal references, word choice, and syntax. Table 1 below presents the major differences in these five aspects when converting the written language in the magazine to the oral language for the video script.

For *formality*, the sentence describing the Pine Garden was originally phrased formally with a logical explanation that included the geographical information about the garden; but in oral language, it was marked by an informal conjunction "so," which initiated the introduction to the garden. As an introductory video, the *depth* of language may be compromised for an easier listening experience, allowing the visual presentation in the video to present how the landscape looked like. *Personal reference* was also one feature often used in films or videos. The written text that introduced the Yanzikou trail was a plain, objective description and the oral form using the personal pronouns as "we" and "you" made it more conversational and interactive. *Word choice* was an important indicator of the genre for a specific purpose. The statement describing the Amis people's different festivals contained less common words "glamorous" and "celebration" and a long phrase "a variety of." In the oral form, the description was simpler with fact-based language, corresponding to the target audience's feedback. Lastly, sentences with complex *syntax* were considered a hindrance to comprehension and the revision was aimed to make sentences syntactically simpler and more straightforward, as shown in Table 1.

Table 1. Differences between written and oral language

Difference	Written language (magazine content)	Oral language (video script)
1. Formality	More formal Pine Garden: "The Pine Garden lies nearby in the plain and the south of Melian District at the north-eastern corner of Hualien city…"	Less formal Pine Garden: "**So**, **we** are going to learn about the Japanese colonial history in Hualien. The Pine Garden lies in the north-eastern corner of Hualien city…"
2. Depth	Greater precision and detail Qingshui Cliff: "Qingshui Cliff is a coastal cliff above sea level in Xiulin township, Hualien county. It is 21 kilometers in length and rises 24 kilometers from the Pacific Ocean…. The best view of this can be observed from the Suhua Highway, which crawls below the magnificent vertical cliff."	Less detailed descriptions Qingshui Cliff: "Hey! Look at the steep cliff and the clean water. Here is the Qingshui Cliff. Now we are on the platform of the Suhua Highway, and the best view of the cliff can be observed from here."
3. Personal reference	Fewer personal references (i.e., pronouns of the first- and second-person singular and plural) Yanzikou Trail: "The 1.4km long Yanzikou trail offers another equally heart-stopping experience."	More personal references (i.e., pronouns of the first- and second-person singular and plural) Yanzikou Trail: "Now we are on the Yanzikou trail. In case of falling rocks, you will need this. Now you are safe. Let's go!"
4. Word choice	Longer and less common words The Amis: "The Amis hold a variety of glamorous festivals for celebration. These festivals include Sea Festival, also known as Catching-Fish Festival, Harvest Festival, and other minor festivals."	Shorter and more common words The Amis: "They hold various festivals, such as Catching-Fish Festival, Harvest Festival, Water-Fetching Festival, etc."
5. Syntax	Syntactically more complex Taroko Gorge: "Each passage informed me that the gorge was a natural wonder, which has captivated the human race since its discovery."	Syntactically simpler Taroko Gorge: "The gorge has been a natural wonder since its discovery."

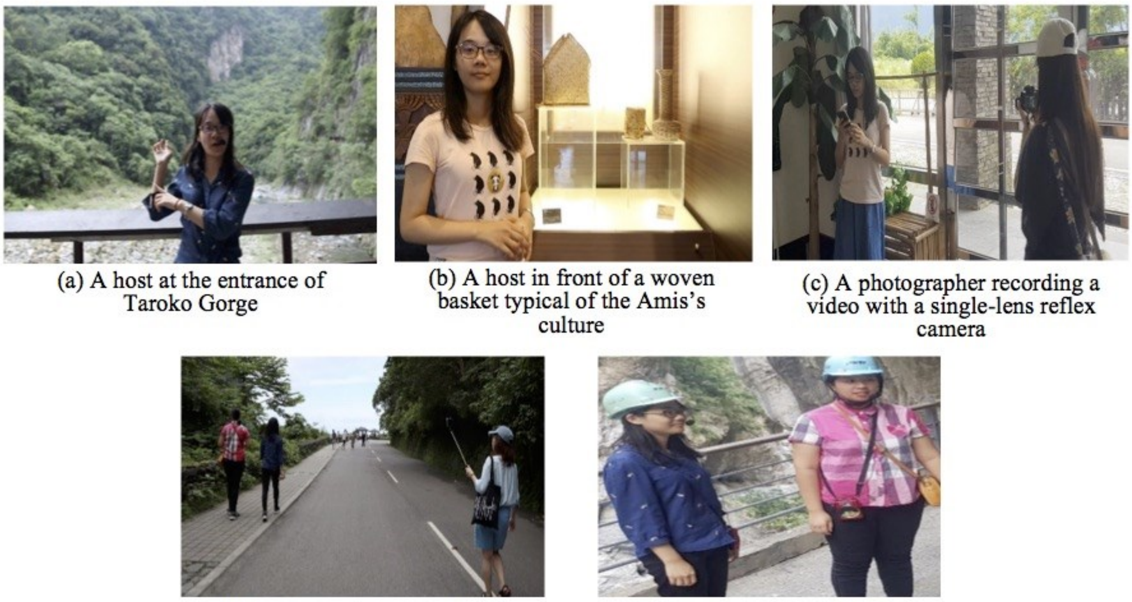
4.1.3. Stage III: Creating multimodal materials for composing the video

In this section, the analysis focused on what role the target audience played, how different modes were orchestrated, and most essentially, how the two interplayed during this stage. Our analysis showed that the multimodality framework enabled the research team to pay attention to a chosen mode while assembling other modes to strengthen the given mode.

Linguistic mode. The written discourse in the magazine has been extensively shifted to an oral genre for video scripts in relation to formality, depth, personal references, word choice, and syntax. However, the oral scripts would not be effectively presented without taking *the aural mode* into play. To familiarize themselves with the oral language during filming, the two hosts had to rehearse the oral script several times to make sure they could enunciate the words clearly and that they were familiar with their lines. Also, *to engage their target audience*, the two hosts added some impromptu phrases to make their speech more natural, for example, adding the impromptu sentence "Here is the Qingshui Cliff" to signal to the audience that the introduction to Qingshui Cliff was about to start.

Visual mode. To create the most appropriate multimodal materials for the most effective visual experience for the target audience, the research team assembled different modes to achieve the communicative purpose. For example, linguistic mode was added to attract our target audience's attention when talking about Taroko Gorge; one of the hosts opened the introduction by announcing, "Hi! We are on the Shakadang Trail, part of the Taroko Gorge. Let's go inside for an adventure." (Figure 1a). Another example is taking *spatial mode* into consideration to enhance the best visual experience. For instance, when talking about the Amis's weaving techniques, one of the hosts stood in front of a weaving handicraft, and the camera captured the handicraft so that *the audience* could view the handicraft during the introduction (Figure 1b).

Aural mode. The aural mode was closely interrelated to the *visual mode* in the composing process because the more sources that the visual mode could capture, the more aural input there was to enrich the aural experience during the output stage (see Figure 1c, 1d, & 1e). For example, using both a camera and a smartphone not only expanded the visual experience but also the aural sources. In addition, to ensure that the *target audience* could experience the oral script (the aural mode) provided by the two hosts, they used voice amplifiers. Amplifiers were especially important when the surrounding was noisy, such as cicadas chirping in the trees (Figure 1e), as experienced in the Pine Garden. This also explains why the hosts did not begin videotaping until all people and cars had left the area.



(a) A host at the entrance of Taroko Gorge

(b) A host in front of a woven basket typical of the Amis's culture

(c) A photographer recording a video with a single-lens reflex camera

(d) A photographer recording the two hosts taking a walk with a smartphone

(e). Two hosts carrying voice amplifiers

Figure 1. Multimodal materials including linguistic, visual and aural modes

Gestural mode. The target audience desired to see hosts in the film. Therefore, facial expressions, gestures, and movements became an essential part of the meaning-making process. For example, having the hosts to walk along the trail and to have a casual conversation was to foster a relaxing feeling, which once again was strengthened via an assembly of four other multimodal modes to achieve this communicative purpose. As always, to engage the target audience, the hosts made good use of gestures and facial expressions corresponding to the ambience set for the filming. For example, at the Ji-An Shrine, one of the hosts signaled the numbers 1, 2, 3 with her fingers (Figure 2a) to emphasize the three functions of the shrine to be introduced and smiled to demonstrate the amiability portrayed throughout the filming.

(a) A host speaking with gestures in a hall inside Ji-An Shrine

(b) A host reciting a poem about Taroko Gorge in the background

(c) A host standing on the right, instead of in the middle

Figure 2. Multimodal materials including linguistic, gestural, and spatial modes

Spatial mode. How the hosts position themselves on the screen in the filming process was the central issue informed by the spatial mode. At times, they decided to leave the whole screen/space to the audience. For example, when reciting a poem about Taroko Gorge, the photographer captured the natural scenery only and did not include the host so that the poem could be put on the screen (linguistic mode) with the natural scenery as the backdrop (visual mode) and the recitation (aural mode) as the background voice (Figure 2b). Even when the host was part of the screen, the major focus was on *the audience*. For example, when introducing the distinctive foods of the Amis, the host stood on the right instead of in the middle to make sure that the foods could be best viewed by the audience (Figure 2c).

4.1.4. Stage IV: (Re-)composing the two versions of the video

When the first version of our video was produced, an internal review by the research team found that three aspects of the video needed to be improved. First, the pace was too slow. Second, only one song was used for background music and it became monotonous; third, the color and size of the subtitles made reading difficult. Although the research team had the ability to self-critique the self-made video, they did not have the enough digital composing skills to resolve the problems identified. Therefore, they decided to turn to a video-editing professional who was well informed by the research team about the preferences of our target audience and the communicative purpose of the video to help create a second version of the video in order to eliminate the flaws identified.

Compared with the first version, the differences in the second version include using a new editing software to ensure the quality, maintaining the overall style throughout the film, changing the font of the subtitles, inserting transitional clips to different scenes, creating ear-catching sound effects, adding dazzling special effects, and adjusting the saturation and exposure of the output (see Table 2).

Table 2. Differences between the two versions of our video

Aspect	Version 1	Version 2
Editing software	PowerDirector 365	Final Cut Pro X
Overall style (visual & spatial mode)	A documentary coupled with funny elements at the END of the video clip	A documentary integrated with funny elements THROUGHOUT the video clip
Subtitle (visual mode)	English (font: Times New Roman) and Chinese (font: DFKai-SB) subtitles	English and Chinese subtitles (font: GenJyuuGothic)
Background music (soundtrack) (aural mode)	One song through the video	Seven different songs in the video (to mark different sections)
Transition (visual mode)	Not well-defined	Clean-cut
Sound effects (aural mode)	No	Yes
Functional special effects (visual mode)	Simple	Dazzling
Ornamental special effects (visual mode)	No	Yes
Color saturation (visual mode)	No adjustment	Adjustment of color saturation and exposure

4.2. RQ2: What can our target audience's responses to the videos inform us about the orchestration of multiple modes when transforming a magazine to a video?

The analysis of the target audience's responses to the two versions of the video informed us of three major insights into the orchestration of multimodal modes involving a target audience. First, the results show that content and format were interdependent during the orchestration process, greatly enhancing the effectiveness of the product. Second, the target audience's feedback served as critical information for the video creators to reconsider their multimodal meaning-making process. Finally, interdisciplinary collaboration with a professional manifested a new possibility for experiencing multiliteracies.

4.2.1. The content and the format as mutually conducive

Most of the interviewees rated the content of the second version (*M* = 7.8) much higher than the content in the first version (*M* = 6.4) even though the content about the six scenic spots was all generated from the first version (see Table 3) Why so? A short introductory video clip with funny and interactive elements incorporated with mini no-good (NG) clips was inserted before each scenic spot was presented while the NG clips could not be seen until the very last part of the video in the first version. When Interviewee 5 was asked why she preferred the Ji-An Shrine in the second version (*M* = 8) to that in the first version (*M* = 6), she responded that "the special effects and sound effects" of the second version and "the NG clips inserted throughout the video" made the content "more attractive" and "more appealing." Likewise, the interviewees liked the hosting style of the second version (*M* = 7.7) much more than that of the first version (*M* = 5.6), although the two hosts had exactly the same scripts, facial expressions, and movements. Interviewee 2 indicated that the hosting style in the second version was "a lot better" and "more relaxing thanks to the great editing." As indicated, the re-composing of the content through a different orchestration of multiple modes resulted in a better perception of the hosting style to be "more natural and appealing," (Interviewee 3) and "more interesting and attractive" (Interviewee 8).

Table 3. Target audience's responses to the videos

Item		First version (*M*)	Second version (*M*)
WHAT (content)	Qingshui Cliff	6.3	7.5
	Shakadang Trail	6.4	7.7
	Yanzikou Trail	6.3	7.6
	Pine Garden	6.0	7.6
	Ji-An Shrine	6.6	8.0
	Hualien Indigenous Tribe Museum	6.6	8.2
HOW (format)	Overall style	6.4	7.5
	Length	6.1	7.8
	Hosting style	5.6	7.7
	Subtitles	5.6	8.0
	Background music	5.5	8.7
	Sound effects	4.7	8.2
	Special effects	5.2	7.9

4.2.2. Gaining feedback from the target audience as renegotiation of meaning-making

When the target audience was first asked whether or not they would like to include special effects and sound effects as part of the video composing, the "no special effects" option won out (66%) in comparison with other choices provided, and 80% of them did not want sound effects. The major reason was that they were afraid that the effects might distract the viewers' attention to the content of the video. Interestingly, when the target audience was asked to express their preferences for the two versions they rated the second version (*M* = 8.2) (with added special and sound effects) noticeably higher than the first version (*M* = 4.7) featuring no effects based on the results of the survey. The special effects made the second version "fun and not too serious," (Interviewee 2), "funny and more entertaining," (Interviewee 5), and lent more "visual appeal" (Interviewee 6). The possible explanation for this discrepancy may be that the majority of the target audience might have expressed their preferences based on their previous experience of special and sound effects when viewing videos. Without the chance to compare and evaluate the two versions of the video, the target audience would not have learned that appropriate special and sound effects would not distract viewers' attention. Rather, it would enhance the interactivity between the viewers and the video, leading to a better engagement and viewing experience. Likewise, the research team would not have been able to learn that meaning-making was a recursive process of re-negotiation and re-adaptation had it not been for the involvement of the target audience.

4.2.3. Interdisciplinary collaboration as new possibilities for learning

The responses of the target audience highlighted the importance of interdisciplinary collaboration. Due to the insufficient digital composing skills, the composing process would have ended after the research team made their first trial had it not been for the collaboration with a professional with an expertise in video composing. Without the collaboration, the research team probably would not have been able to "fix" the problems even when they had the ability to identify the flaws of their own video composing. Second, the collaboration with a professional helped the research team realize how content could be received by the target audience differently with different multimodal composing skills. The research team would not have known how many levels/possibilities of meaning-making could be achieved through assembling different multimodal modes and how their target audience would respond differently to the two versions of the video with different orchestration. The learning would not have been possible without the collaboration with a professional with the relevant expertise required for digital compositing.

5. Discussion

This case study, similar to Hafnar (2014) and Yeh (2018), yielded results leading to a four-stage framework based on the transforming process. It shares three features with the previous studies: composing the video script, filming the video, and editing the video. However, what distinguishes the framework here is the role of a target audience in a multimodal literacy project. A real target audience's involvement in the (re-)composing process enhances the overall effectiveness of the meaning-making and communication outcome. The multimodal orchestration process, in the meantime, provided the EFL learners a *capstone* learning experience beyond language per se, which may be examined by the framework of "*Learning by Design*" proposed by Kalantzis and Cope (2005).

5.1. The role of a target audience in a multimodal literacy project

Previous studies examining multimodal literacies in terms of composing (e.g., Campbell & Cox, 2018; Jiang, 2018) do not seem to specify a target audience, except for Hafner (2014) and Yeh (2018). While both Hafner and Yeh mentioned the target audience in their studies, the specific role(s) the audiences played in the "reading" stage of Hafner's model and the "composing stage" of Yeh's are not clear. In our study the authentic target audience offered their preferences for what they wished to see and for how the video should be made by way of a questionnaire based on the five modes in the multimodality framework and by elaborating on the results of the questionnaire in in-depth interviews. In short, our target audience had a direct impact on the process, as their choices and preferences shaped the content and format of our video.

Prior to the second stage, the research team had been well informed by applied linguistic studies that ample differences between oral and written language do exist (see Stage II); in addition, our target audience as key informants told us what type of oral language style (e.g., simple, fact-based oral language) they wished to see and hear in the video. Hafner (2014) also noticed that different people may draw on different "discursive forms (genres, registers, and styles)" (p. 657) to serve different purposes (e.g., written language in a lab report for an academic audience versus oral language in a Facebook update for friends). However, it was not clear whether the target audience(s) played a role in determining how language should be shaped to suit their interests and needs. Similarly, Yeh (2018) mentioned that her students "[used] English to introduce Taiwanese people, customs, cultural values, or architectural history" (p. 30) to "a global audience" (p. 36). There was not much elaboration on how their English was used in relation to the global audience in mind, e.g., genre, style, word choice. Our study shows a deliberate effort to draw our target audience into the preferred type of oral language for the video script.

Our target audience also served as critical reviewers in the final stage of video creation, providing key insights into the video-editing skills required for effective video production. Without the target audience reviews of the two versions of the video, it would have been difficult to ensure the effectiveness of the video production. Instead of seeking feedback from the target audience, Hafner (2014) asked his students to share their final videos online and obtain feedback from peers after the sharing session, while Yeh (2018) and her research team graded her students' video clips. In contrast, our target audience's ratings on and insights into the differences between the two versions of the video enabled a critical examination that helped understand what and how multimodal elements could be re-orchestrated to expand a viewing experience that our target audience had not expected.

5.2. The knowledge processes in multimodal orchestration

In terms of multimodal orchestration, the studies previously reviewed and the current one all presented an evolving process along which an effective video was produced. However, what made this case study stand out was the evaluation of how such orchestration processes also facilitated the knowledge construction processes conducted with an analytic framework including four elements: *experiencing the known and the new,* *conceptualizing by naming and theorizing, analyzing functionally and critically, and applying appropriately and creatively* (Kalantzis & Cope, 2005).

The students embarked on this project by reviewing the magazine they had produced in an earlier course. The content was about a place of which the people and culture were already known to them. They had read and written about the place incorporating three modes (linguistic, visual and spatial) and presented it to their imaginary target audience in a magazine format as a case of print media. When enrolling in this capstone course, they determined to venture into *the new* by transforming the magazine into a video in order to reach a larger potential population which led to a series of, but not necessarily linear, processes of constructing knowledge.

Bringing their initial understanding and experience on multimodality and the concept of target audience, the students thus began to reach a group of authentic target audiences to collect detailed information for the later scripting, filming, and (re)composing processes. These processes helped students conceptualize and concretize their learned knowledge. First of all, the survey questionnaire was designed based on the five modes that have been theorized and applied widely in earlier research. Second, converting the discourse style from textual to oral language when scripting in terms of formality, depth, personal reference, word choice, and syntax was a new experience of practicing pragmatics for a specific target audience. This practice broadened and deepened their language learning in a new context. Next, in addition to applying the knowledge about language, multimodality, and target audience innovatively, they also experienced an interdisciplinary collaboration with a professional in filming which greatly expanded their perception of how the various multimodal semiotic resources should be taken into consideration when composing and recomposing the video.

Finally, the re-composition process was conducted due to a critical internal analysis by the students themselves with the help of a professional and an evaluative comparison between the two versions of the video given by the target audience. For our students, their multimodal orchestration competence built along the project enabled them to self-criticize the quality of the video, hence an improved version for the video was recomposed. For the target audience, they were able to reconsider their original requests (no sound and special effects) and accepted the new video as a better version. The new feedback given by the target audience, in turn, resulted in a new discernment in our students in that the target audience's preferences should be flexibly adapted and adjusted.

6. Conclusion and implications

Being situated in the 21st multimodal global society, literacy educators are confronted with the demand to equip their students not only with fundamental literacy skills (reading and writing) but also with multimodal literacies in order to achieve communicative purposes effectively. Hence, this case study investigated how a printed text (i.e., a magazine with still text-image compositions), as a common mode in the majority of EFL classrooms, can be transformed into a multimodal one (i.e., a video with moving text-image compositions) when involving a specified target audience. The results show that the transformation from a paper format to a digital video can be enhanced through the four-stage digital composing process, and the orchestration of multiple modes in a video can be best evaluated via the target audience's authentic comments and feedback. The findings also reveal that the EFL learners, as composers of the video in this case study, whose knowledge about digital composing process and multiliteracies could be (re-)conceptualized, expanded, and renewed when applying the multimodality framework. The whole process reported in this case study, thus, exemplified the knowledge processes proposed by Kalantzis and Cope (2005) and this *Learning by Design* framework also helped us examine and confirm such multimodal practices indeed facilitate knowledge construction and uplift a multiliteracies capacity in EFL learners. Three important implications for multimodal literacies are generated based on the major findings and discussion.

First, from a theoretical perspective, the results indicate that a well-rounded video composing process should not overlook the critical roles that a target audience can play throughout the different stages of the process. Thus, it implies that a target audience should be specified and integrated into the multiliteracies curriculum so that learners can work with the target audience to determine what to include, how to produce a video, and to what extent their video production is effective from the target audience's perspective. Second, the target audience's significantly different ratings on the two different versions of the video highlight the importance of multimodal competencies and skills. As language majors, who tend to have a better control of the linguistic mode, may not have the capabilities to integrate other modes. A multiliteracies curriculum should create a mechanism for interdisciplinary collaboration, for example, allowing design majors to take multiliteracies courses with language majors to facilitate such collaboration and to enhance the multimodal process and product. Lastly, the study indicates that the knowledge and skills to combine, orchestrate, or mix different modes should not be viewed as a given even with a curriculum component featuring interdisciplinary collaboration. Ample multimodal learning opportunities should be designed and provided to engage EFL learners in multiliteracies activities so as to equip them with knowledge and competences to experience, conceptualize, analyze, and apply from the known to create new possibilities for multiliteracies education.

7. Limitations and further studies

This study was limited in three ways. First, the knowledge processes would have been more complete if a third version of the video had been created based on the target audience's ratings and feedback on the two versions of the video. Further research can investigate whether or not a third version of the video can provide an even more satisfying viewing experience and create more in-depth knowledge processes. Second, the professional invited did not involve in the four-stage transforming process. It will be interesting to examine if a different transforming process would emerge when a professional is invited to participate in the very beginning of the project, or when the research team is involved in a different classroom structure, such as working with class/teammates with multimodal expertise. Lastly, the four-stage model was developed based on a group of international students designated as the specified target audience in our study. The preferences for the five modes may be limited to the perspective of this specified target audience. It is therefore suggested that more than one group of target audience be involved in a case study to examine whether or not different preferences among different groups of target audience can be observed and how the differences impact each stage of the video composing process.

Acknowledgement

The authors would like to thank the anonymous reviewers for their constructive comments which helped improve this paper to a great extent. This project is supported by the Ministry of Science and Technology, Taiwan, under the grant nos. MOST 109-2511-H-003-026, MOST 106-2511-S-003 -015 -MY3, MOST 108-2410-H-011-017, and MOST 108-2813-C-011-037-H.

References

Alexander, J., & Rhodes, J. (Eds.). (2014). *On multimodality: New media in composition studies*. Urbana, IL: Conference on College Composition and Communication-NCTE.

Campbell, L. O., & Cox, T. D. (2018). Digital video as a personalized learning assignment: A Qualitative study of student authored video using the ICSDR model. *Journal of the Scholarship of Teaching and Learning, 18*(1), 11-24.

Cho, Y., & Choi, I. (2018). Writing from sources: Does audience matter? *Assessing Writing, 37*, 25-38.

Crowe, S., Cresswell, K., Robertson, A., Huby, G., Avery, A., & Sheikh, A. (2011). A Case study approach. *BMC Medical Research Methodology, 11*, 1-9.

DeVito, J. A. (1965). Comprehension factors in oral and written discourse of skilled communicators. *Speech Monographs, 32*, 124-128.

Einhorn, L. (1978). Oral and written style: An Examination of differences. *Southern Speech Communication Journal, 43*(3), 302-311.

Gunel, M., Hand, B., & McDermott, M. A. (2009). Writing for different audiences: Effects on high-school students conceptual understanding of biology. *Learning and Instruction, 19*, 354-367.

Hafner, C. A. (2014). Embedding digital literacies in English language teaching: Students' digital video projects as multimodal ensembles. *TESOL Quarterly, 48*(4), 655-685.

Hung, S.-T. A. (2019). Creating digital stories: EFL learners' engagement, cognitive and metacognitive skills. *Educational Technology & Society, 22*(2), 26–37.

Jewitt, C. (2006). *Technology, literacy, and learning: A Multimodal approach*. London, UK: Routledge.

Jiang, L. (2017). The Affordances of digital multimodal composing for EFL learning. *ELT Journal, 71*(4), 413–422.

Jiang, L. (2018). Digital multimodal composing and investment change in learners' writing in English as a foreign language. *Journal of Second Language Writing, 40*, 60-72.

Kakh, S. Y., Mansor, W. F., & Zakaria, M. H. (2014). Rhetorical analysis tasks to develop audience awareness in thesis writing. *Procedia-Social and Behavioral Sciences, 98*, 806-813.

Kalantzis, M., & Cope, B. (2005). *Learning by design.* Melbourne, Australia: Victorian Schools Innovation Commission in association with Common Ground Publishing.

Kress, G. (2000). Multimodality: Challenges to thinking about language. *TESOL Quarterly, 34*(2), 337-340.

Kress, G. (2003). *Literacy in the new media age*. London, UK: Routledge.

Lin, T.-B., Li, J.-Y., Deng, F., & Lee, L. (2013). Understanding new media literacy: An Explorative theoretical framework. *Educational Technology & Society*, *16*(4), 160–170.

Martinez-Insua, A. E. (2019). Scientific writing and the contentfulness of subject themes. How science was explained to (lay) audiences. *Journal of Pragmatics, 139*, 216-230.

Mills, K. A. (2010). "Filming in progress": New spaces for multimodal design. *Linguistics and Education, 21*, 14-28.

Mills, K. A. (2011). *The Multiliteracies classroom*. New York, NY: Routledge.

Mills, K. A. (2015). Doing digital composition on the social web: Knowledge processes in literacy learning. In B. Cope & M. Kalantzis (Eds.), *A Pedagogy of multiliteracies: Learning by design* (pp. 172-185). London, UK: Palgrave Macmillan.

Monte Mŏr, W. (2015) *Learning by design*: Reconstructing knowledge processes in teaching and learning practices. In B. Cope & M. Kalantzis (Eds.), *A pedagogy of multiliteracies: Learning by design* (pp. 186-209). London: Palgrave Macmillan.

National Council of Teachers of English. (2013). *NCTE Framework for 21st Century Curriculum and Assessment*. Retrieved from https://secure.ncte.org/library/NCTEFiles/Resources/Positions/Framework_21stCent_Curr_Assessment.pdf

New London Group. (1996). A Pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review, 66*(1), 60-92.

Pan, J. H., & Zhang, R. J. (2019). Preliminary investigation of the current situation and influencing factors of international students in Taiwan under the background of new southbound policy. *Taiwan Educational Review, 8*(2), 154-175.

Schallert, D. L., Kleiman, G. M., & Rubin, A. D. (1977). *Analyses of differences between written and oral language* (Report No. 29). Champaign, IL: University of Illinois at Urbana-Champaign.

Strauss, S., Feiz, P., & Xiang, X. (2018). Written versus spoken grammar. *The TESOL Encyclopedia of English Language Teaching, 1*, 1-8.

Wagennar, T. C. (1993). The Capstone course. *Teaching Sociology, 21*(3), 209-14.

Wilkinson, A. M. (1971). *The Foundation of language: Talking and reading in young children*. London, UK: Oxford University Press.

Wong, T. Y. (2005). Writers' mental representations of the intended audience and of the rhetorical purpose for writing and the strategies that they employed when they composed. *System, 33*(1), 29-47.

Yeh, H. C. (2018). Exploring the perceived benefits of the process of multimodal video making in developing multiliteracies. *Language Learning & Technology, 22*(2), 28-37.

Wu, B., Hu, Y., & Wang, M. (2020). How do Head-mounted Displays and Planning Strategy Influence Problem-solving-based Learning in Introductory Electrical Circuit Design? *Educational Technology & Society*, *23* (3), 40–52.

How do Head-mounted Displays and Planning Strategy Influence Problem-solving-based Learning in Introductory Electrical Circuit Design?

Bian Wu[1*], Yiling Hu[1] and Minhong Wang[1,2]
[1]Department of Education Information Technology, East China Normal University, Shanghai, China // [2]KM&EL Lab, Faculty of Education, The University of Hong Kong, Hong Kong, China // bwu@deit.ecnu.edu.cn // ylhu@deit.ecnu.edu.cn // magwang@hku.hk
[*]Corresponding author

(Submitted May 5, 2020; Revised June 19, 2020; Accepted July 29, 2020)

ABSTRACT: The high-fidelity and interactivity afforded by head-mounted displays (HMD) has a potential to improve learning in problem-solving contexts. However, there is a lack of studies with mixed findings on the efficacy of HMD in the development of problem-solving competence. Moreover, the integration of learning strategies with HMD supported learning is often overlooked. This study aimed to address the gap by investigating whether the simulation media (HMD or 2D simulation environments) and learning strategy (with or without planning) may influence student learning with problem-solving tasks. The results show that the HMD and planning groups outperformed other groups in simulated problem-solving tasks, and in transferring the competence to real-world tasks. Students using the HMD perceived a higher level of sense of presence, self-efficacy, and simulator acceptance; but they reported a higher level of mental workload and simulator sickness than those using the 2D simulation. Implications of the findings are also discussed.

Keywords: Virtual reality, Simulations, Learning strategies, Problem solving

1. Introduction

Research on virtual reality in education has proliferated over the past decade. More recently, the head-mounted display (HMD) technology that provides personal perspectives in virtual reality settings offers a sense of immersion during learning, and is therefore being called immersive virtual reality (IVR). By adding immersive and interactive alternatives to traditional classroom settings, HMD environments may facilitate experiential learning, conceptual and procedural understanding, and psychomotor skill development (Concannon, Esmail, & Roberts, 2019). In particular, HDM may benefit learning in problem-solving contexts, which is material-saturated and requires body engagement (Chen, Hong, Sung, & Chang, 2011; Johri & Olds, 2011). HMD technology can enhance problem-solving-based learning by inducing more epistemic actions, i.e., actions that augment human cognition and help reveal hidden information (Jin & Lee, 2019).

Previous studies found that HMD environments of high-fidelity improve users' sense of presence, emotion, motivation, and technology acceptance (Chang, Heo, Yeh, Han, & Li, 2018; Cooper, Park, Nasr, Thong, & Johnson, 2019; Kwon, 2019; Makransky, Borre-Gude, & Mayer, 2019; Makransky & Lilleholt, 2018; Shu, Huang, Chang, & Chen, 2019; Zhang et al., 2020). Nevertheless, research also shows that HMD may increase learners' mental workload and sickness symptoms (Jost, Cobb, & Hämmerle, 2019; Meyer, Omdahl, & Makransky, 2019). It remains unclear if these findings of personal factors can be generalized to different HMD-based learning contexts, because the requirements of immersiveness and mental workload as well as affordances of HMD regarding visualization and interactivity may vary for different learning goals and in different VR-based learning applications (Wu, Yu, & Gu, 2020). In addition to self-reported learning experience, a few studies explored the effects of HMD on improving learning outcomes and found promising results regarding learning efficiency, skill acquisition and conceptual understanding (Huang, Luo, Yang, Lu, & Chen, 2020; Jung & Ahn, 2018; Parmar et al., 2016). While HMDs have a potential to foster experiential and problem-solving-based learning, there are, however, inadequate studies investigating the effects of HMD on problem-solving competence and the use of hands-on activities for performance-based assessment (Lamb, Antonenko, Etopio, & Seccia, 2018). In addition, existing studies on the effect of HMD on learning performance have not considered how well the competence was transferred from the simulated environment to real situations (Yang et al., 2018; Falloon, 2020).

Compared to computer-based 2D simulations or desktop VR, HMDs provide a high level of fidelity. However, prior research that compared HMDs and desktop simulations regarding their effects on problem-solving-based learning reported mixed results (Makransky, Terkildsen, & Mayer, 2019; Parmar et al., 2016). To optimize VR-supported simulation learning, Fowler (2015) and Meyer et al. (2019) suggested that VR as an innovative media should be integrated with theoretically sound pedagogical design. For example, Parong and Mayer (2018) found that students who summarized the lesson after each HMD segment performed significantly better on the post-test of conceptual understanding than the groups that did not, and Meyer et al. (2019) found that pre-training and HMD had an interaction effect on learning. Unterrainer and Owen (2006) indicated that participants who made a full mental plan before movements performed better in solving problems than those who immediately began task-related movements. Thinking aloud at the beginning of the process can stimulate conceptual knowledge understanding, which helps to improve the ability to devise a plan to solve the problems (Kani & Shahrill, 2015). Therefore, it is of interest to investigate whether a planning strategy can improve students' problem-solving performance in HMD environments.

Specifically, as learning electrical circuit design is concerned, previous studies (Zacharia, 2007; Chen et al., 2011; Jaakkola, Nurmi, & Veermans, 2011) revealed the benefits of combining simulation-based learning with real experimentation. However, although desktop simulation software is widely used in science and engineering education, it seems a challenge to transfer knowledge from 2D-based virtual environment to real practice (Richard & Taylor, 2015). Therefore, the study extends the line of inquiry to explore the effectiveness of HMD-based simulation context in support of problem-solving learning in electrical circuit design, which in turn, can contribute to our understanding regarding whether and how HMD can better complement the traditional approaches of lecturing and real lab experiment.

In sum, existing studies have reported positive evidence of the utility of HMD technology in educational settings. However, these studies have placed more attention to learner perceptions than learning outcomes. While a few studies explored the effects of HMD on improving knowledge retention and skill acquisition, and the effect of HMD on problem-solving performance remains unclear, let alone the effects of HMD on the transfer of learning beyond the simulation settings. Therefore, this study aimed to investigate whether HMD and planning strategies would influence student learning with problem-solving tasks in introductory electrical circuit design. The following three research questions were addressed.

RQ1: Will the simulation media (i.e., HMD or 2D simulation environments) and the learning strategy (i.e., with or without planning) influence students' problem-solving performance in simulated environments of electrical circuit design?

RQ2: Will the simulation media (i.e., HMD or 2D simulation environments) and the learning strategy (i.e., with or without planning) influence students' transfer of problem-solving competence from simulated tasks to real-world tasks in electrical circuit design?

RQ3: Will the simulation media (i.e., HMD or 2D simulation environments) and the learning strategy (i.e., with or without planning) influence students' perceptions of the problem-solving-based learning experience in electrical circuit design?

The perceptions include the participants' self-efficacy, sense of presence, simulator acceptance, perceived mental workload, and perceived simulator sickness, and their opinions on the use of HMD in problem-solving-based learning.

2. Method

2.1. Context and participants

Fifty-two first- and second-year undergraduate students (21 males, 31 females, mean age = 18.4 years, SD = 0.75 years) participated in this study voluntarily. They were teacher-students enrolled in an educational technology program at a university in eastern China. They participated in this study for the learning of electrical circuit design, as part of the professional development for pre-service STEM teachers. The exclusion criterion was the enrolled participants had taken any introductory electrical circuit design course at university level before this study. The study employed a two-factor experimental design: simulation media (HMD or 2D simulation environments) and learning strategy (with or without planning). The participants were randomly assigned to four conditions, with thirteen students in each condition. This research was approved by the ethical review board of the researchers' institution. All participants signed informed consent forms for their participation in this study.

Electrical-circuit design involves complex skills. Although real lab experiments or using simulators seems to be norm in teaching circuit design, there are many factors influence its adoption in China, such as higher education instructors’ lack of student-centered teaching competence, knowledge-based and exam-oriented curriculum planning, and limited time for practice teaching (Wu, Liu, & Yi, 2012). The dominant pedagogy of this learning subject remains “chalk and talk,” which limits students’ practical experience (Mills & Treagust, 2003). In this study, students were asked to learn by using various electrical elements and devices to develop artifacts (i.e., electrical circuits) that meet the given requirements and constraints. In a simulation environment, if the electrical elements were connected correctly, the circuits would function as expected.

2.2. Learning environments and materials

For the HMD condition, we used Short Circuit VR, which is a free electrical VR simulator developed by Bauwens and Ho (2018). The simulation was displayed through HTC VIVE with a 110° horizontal field of view, a resolution of 2160 × 1200 pixels, 90 fps, 6-degree-of-freedom, handheld controllers and on HP Alienware (CPU i7-9700K 3.6 GHz, RAM 16.0 GB, GPU NVIDIA GTX1660Ti). For the desktop 2D-simulation condition, we used Breadboard Simulator v1.0, which is open-source software developed by Shah (2018). The two simulation environments for circuit design are shown in Figure 1.

We provided two circuit-design problems for the simulated learning experiment, which were conducted on either a 2D simulation on a PC or on an HMD-based simulation environment. The first problem required students to design a circuit to light up an LED. The students needed to understand the working principle of diode, the polarity of the LED, apply Ohm’s law to choose the correct resistance, and make sure the current flowing through the LED was within the safe range. The second problem required students to control a seven-segment display such that the number “1” could be turned on and off using a push button. To complete this circuit design, the students needed to understand the mechanism and usage of some basic electrical components. The second problem was more difficult than the first one. The problems were developed by an instructor of introductory electrical circuit design course to make sure the experimental materials were comparable to the tasks that are normally given in the course.

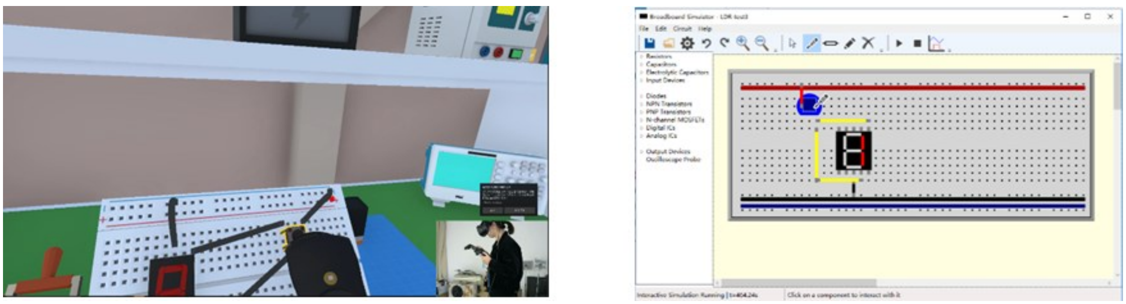


Figure 1. Screenshots from the simulation environments for electrical circuit design: The HMD condition using a Short Circuit VR environment (left) and the 2D simulation condition using a Breadboard Simulator environment (right)

2.3. Instruments

The pre-test questionnaire collected the participants’ demographic information including their gender, age and whether they have taken any introductory electrical circuit design courses at university level or not. The post-test questionnaire consisted of measures of sense of presence, self-efficacy, mental workload, simulator sickness, and simulator acceptance (see Appendix 1 for the complete questionnaire). The sense of presence scale (four items) was adapted from Witmer and Singer (1998). The self-efficacy of learning scale (three items) was adapted from Meluso, Zheng, Spires, and Lester (2012). The mental workload scale (four items) was adapted from Hart and Staveland (1988). The simulator acceptance scale (five items) was adapted from Davis (1989) and Venkatesh and Bala (2008). The simulator sickness scale (three items) was adapted from Kennedy, Lane, Berbaum, and Lilienthal (1993). Cronbach’s α was used to evaluate the internal consistency of each dimension. The Cronbach’s α values was 0.8 for simulator acceptance, 0.81 for sense of presence, 0.72 for self-efficacy, 0.9 for mental workload and 0.58 for simulator sickness. With the exception of the value for simulator sickness, all these values were above the generally accepted cut-off for a satisfactory level of reliability (> 0.7) (Nunnally, 1978). The low reliability for simulator sickness was related to the participants’ relative discomfort, such as their experience of dizziness, fatigue, or blurred vision. In addition to these scales, the 26 participants in the two HMD groups were asked two open-ended questions about whether they preferred learning problem-solving via HMD or via conventional learning methods (i.e., lecturing plus lab-based practice), and the pros and cons of using HMD to support problem-solving-based learning.

Problem-solving performance was measured in terms of success rate and completion time. A dichotomous variable was used as success rate. This variable was given a value of one if the participants solved the problem independently and successfully. If participants asked for help to solve a problem or failed to solve a problem, the variable was given a value of zero. The participants were informed that their time on task was measured as a proxy for performance. They were given five minutes for warming up and being familiar with their corresponding simulation environments before beginning the problem-solving tasks. The completion time was recorded in minutes.

Transfer of learning can be categorized into near transfer and far transfer (Nokes-Malach & Richey, 2015). The former is called for when students encounter problems that are very similar to the problems they have worked on during the learning phase. The latter is called for when students encounter problems that are new to them in both content and context. To assess the extent to which problem-solving competence was transferred from simulations to real-world tasks, the study included a post-transfer test comprising two further electrical circuit design problems that used a physical breadboard and real electrical components. The first, near-transfer problem asked the participants to control a seven-segment display so that number "1" turned on and off, but using a photoresistor instead of the push button used in the second learning task. This near-transfer problem is very similar to the training problem 2. The second, far-transfer problem was more complex than the two training problems and required the students to use a photoresistor to control a seven-segment display that switched between displaying the number "1" and the number "7." Figure 2 shows the description of the second transfer problem and the real breadboard circuit design.

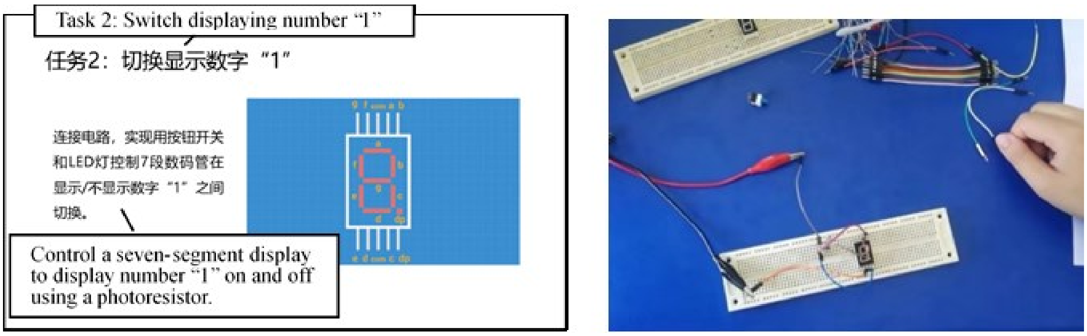


Figure 2. Description of the second transfer problem (left) and the real breadboard-based electrical circuit design (right)

2.4. Procedures

The participants registered the time slot for the experiment and attended the experiment one at a time in the VR-based learning laboratory with the help of one experimenter. The participants first read and signed a consent form. Then, they completed a pre-test questionnaire survey that collected their demographic information. The experimenter provided slides to illustrate the prerequisite concepts for understanding electrical circuits. All participants received the same slides about domain knowledge that are necessary for solving the following problems and learned at their own pace. They could also ask for explanation about the contents of these slides from the experimenter during individual learning. Then, the participants were shown another few slides demonstrated either the desktop 2D simulation environment or HMD and were given adequate time for familiarizing themselves with the assigned simulation environments. The total time for experiment preparation was around 20 minutes.

After warming up in simulation environment to reduce novel effect, the participants were given 60 minutes for the whole experiment of completing the two problems. The four groups were given slides tailored to the four different experimental conditions to describe the first problem with or without the planning requirement. The planning strategy asked the participants to think aloud, with the help of drawings, when planning their electrical-circuit design solutions, before performing the tasks in the simulation environments. They were informed about the time when they reached 30 minutes, but would not be interrupted even if they didn't finish the first problem. When they thought they had completed the first problem, they could move on to read slides describing the second problem and tried to solve it in the same condition as the first one. When they reached 60 minutes, they would be stopped for the experiment no matter they have completed the two problems or not. The experimenter observed and recorded participants' performance in terms of completion time and problem-solving outcome. Webcam and screen capture software were also used to record both participants' gesture and body movement in the real world and their problem-solving process within the simulation environment. The recorded video data was used to verify the observation record.

After completing the experiments, the participants were required to solve two authentic transfer problems. Finally, they completed the post-test questionnaire survey. The post-simulation transfer test and questionnaire survey didn't have a time limit, but all the participants completed around 30 minutes. The research procedures are shown in Figure 3.

Pre-test questionnaire
Prerequisite knowledge illustration (20 min)

Group A (N = 13)	Group B (N = 13)	Group C (N = 13)	Group D (N = 13)
Planning + HMD	HMD	Planning + 2D	2D
Two training problems (60 min)	Two training problems (60 min)	Two training problems (60 min)	Two training problems (60 min)

Post-simulation transfer test (two problems)
Post-test questionnaire
(30 min)

Figure 3. Procedure

2.5. Data analysis methods

To answer the first and the second research questions, we used the statistics of the problem-solving completion time and problem-solving success frequency in the four conditions. A two-way analysis of variance (ANOVA) was performed on the different problem-solving completion times to investigate the variance associated with the different media and the planning strategy. In addition, omega squared (ω^2) was used as an effect-size measure (Howell, 2006). Loglinear analysis was conducted on problem-solving success, media intervention, and planning intervention to examine the association between these three variables. We began with the saturated model, and removed the higher-order interactions first to identify the significant interaction effect. Then, to interpret the interaction, a chi-square test was performed and the odds ratio (*OR*) was calculated as a measure of the effect size.

To answer the third research question, we used descriptive statistics and a two-way ANOVA of sense of presence, self-efficacy, simulator sickness, and simulator acceptance. RStudio v1.2 (RStudio, 2019) and R v.3.6.1 (R Core Team, 2019) were used to run the quantitative analyses. The qualitative data collected through the open-ended questions were evaluated using a thematic analysis approach to triangulate the quantitative findings (Miles & Huberman, 1994). The responses were grouped into themes that reflected respondents' experiences of problem-solving-based learning through HMD. The themes were then ranked in order of frequency, to measure the relative importance of each theme. For each viewpoint, sample responses are presented in the results section.

3. Results

3.1. Problem-solving efficiency

Table 1 shows that the more difficult problem required more time to complete in both the training session and transfer test. Further, the results revealed that the HMD groups completed all four problems faster than did the 2D simulation groups. Moreover, the planning groups completed most of the problems faster than the groups without planning. It is intriguing to see that the group of 2D simulation without planning had a larger variation (SD = 11.76) in completion time of solving the second training problem. This might be due to the second training problem is relatively more difficult than the first one. Another noteworthy finding was the 2D simulation groups almost doubled the completion time in solving two transfer problems compared with the group of HMD with planning. A possible reason is the participants of 2D groups needed more time for processing information related with knowledge transfer into real context.

The two-way ANOVA revealed that the completion time of Problem 1 was the same for all four groups, with no significant main effects of media and method intervention or of the interaction effect. There was a significant main effect of method intervention, $F(1,48) = 5.97$, $p = .018$, $\omega^2 = 0.09$, on the completion time of Problem 2, but no significant main effect of media intervention or interaction effect. The completion time of the near-transfer problem had significant main effects from both media intervention, $F(1, 48) = 6.80$, $p = .012$, $\omega^2 = 0.09$ and method intervention, $F(1, 48) = 4.18$, $p = .046$, $\omega^2 = 0.05$, but no interaction effect. The completion time of the far-transfer problem had a significant main effect from media intervention, $F(1,48) = 5.51$, $p = .023$, $\omega^2 = 0.08$, but not from method intervention or interaction effect. According to Field, Miles, and Field (2013), a ω^2 less than 0.06 indicates a small effect size, between 0.06 and 0.14 is a medium effect size, and greater 0.14 is large effect size. Therefore, all the significant main effects for problem-solving efficiency had small-to-medium effect sizes.

Table 1. Means and standard deviations for problem-solving completion time (in minutes)

	Media						
	HMD		2D simulation		Media	Method	Interaction
	With planning	Without planning	With planning	Without planning	*p* value	*p* value	*p* value
Training problem 1	7.88 (5.37)	7.69 (4.55)	8.80 (6.82)	9.61 (5.92)	.375	.847	.754
Training problem 2	14.23 (8.33)	19.62 (7.87)	14.86 (9.18)	22.23 (11.76)	.536	.018*	.705
Near transfer problem	4.77 (1.74)	8.77 (4.30)	9.38 (5.39)	9.85 (3.36)	.012*	.046*	.112
Far transfer problem	5.69 (3.38)	8.38 (4.87)	10.38 (6.44)	11.08 (7.22)	.023*	.287	.527

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

3.2. Problem-solving performance

Table 2 reports the success/failure frequencies of problem solving for the four problems in the four conditions, and the results of the chi-square test of the association between media or method and performance. The HMD with planning group and the 2D simulation without planning group had the best and worst performance, respectively, in all four problems. The HMD groups had more successes than the 2D simulation groups in all four problems, and the planning groups outperformed the non-planning groups in all problems except the first training problem.

Table 2. Success/Fail frequency of problem solving

	Media				Media x performance	Method x performance
	HMD		2D simulation			
	With planning	Without planning	With planning	Without planning	*p* value	*p* value
Training Problem 1	11/2	10/3	7/6	6/7	.020*	.560
Training Problem 2	8/5	3/10	3/10	3/10	.139	.139
Near-transfer problem	13/0	7/6	8/5	3/10	.011*	.002**
Far-transfer problem	11/2	9/4	8/5	6/7	.080	.244

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

The three-way loglinear analysis using backward elimination produced a final model that retained only one two-way interaction effect, which was the type of media and Problem 1 success, $\chi^2(1) = 5.44$, $p = .020$. Thus, based on the *OR*, the odds of Problem 1 success were 4.08 (1.06, 18.26) times higher for students trained with HMD than for students trained with 2D simulations. In the transfer test, there was a significant association between the type of media and whether the near-transfer problem was successfully solved ($\chi^2(1) = 6.47$, $p = .011$), and between adopting a planning strategy and successfully solving the near-transfer problem ($\chi^2(1) = 9.67$, $p = .002$). Thus, based on the *OR*, the odds of near-transfer problem success were 4.40 (1.19, 18.27) times higher for students trained with HMD than for students trained with 2D simulations, and the success rates were 6.45 (1.67, 29.44) times higher for students who adopted planning than for students who did not plan. Chen, Cohen, and Chen (2010) suggested 1.68 (small), 3.47 (medium), and 6.71 (large) as cutoffs for interpreting the size of the *OR*. Therefore, the significant interaction effects had medium sized effects on problem-solving performance. However, we found no other association between the type of media or planning and problem-solving success in the second training problem or in the far-transfer problem.

3.3. Questionnaire survey results

Table 3 reports the means and standard deviations for sense of presence, self-efficacy, perceived simulator acceptance, mental workload, and simulator sickness. The two HMD groups ranked higher on the five questionnaire constructs than their 2D simulation counterparts. The two-way ANOVA results further confirmed that, compared with the 2D simulation groups, the HMD groups had a significantly higher sense of presence, $F(1, 48) = 19.58$, $p < .01$, $\omega^2 = 0.27$, higher self-efficacy, $F(1, 48) = 6.23$, $p = .016$, $\omega^2 = 0.09$, higher perceived simulator acceptance, $F(1, 48) = 6.48$, $p = .014$, $\omega^2 = 0.10$, but also a heavier mental workload, $F(1, 48) = 6.47$, $p = .014$, $\omega^2 = 0.10$ and more simulator sickness, $F(1, 48) = 8.16$, $p = .006$, $\omega^2 = 0.12$. Therefore, all the significant main effects on the subjective non-cognitive variables showed medium-to-large sized effects. In contrast, there was no significant difference between the groups with and without planning intervention or any interaction effect for all five questionnaire constructs.

Table 3. Means and standard deviations for the questionnaire survey results

	Media						
	HMD		2D simulation		Media	Method	Interaction
	With planning	Without planning	With planning	Without planning	*p* value	*p* value	*p* value
Sense of presence	4.20 (0.50)	4.22 (0.62)	3.40 (0.70)	3.49 (0.65)	$< .01^{**}$	.756	.824
Self-efficacy	4.08 (0.67)	3.64 (0.57)	3.49 (0.88)	3.23 (0.74)	$.016^{*}$	.090	.656
Perceived simulator acceptance	4.22 (0.44)	4.28 (0.59)	3.86 (0.55)	3.88 (0.54)	$.014^{*}$	.796	.877
Mental workload	3.56 (0.79)	3.21 (0.80)	2.87 (1.05)	2.51 (1.22)	$.014^{*}$	.193	1.000
Simulator sickness	2.28 (0.57)	2.36 (0.84)	1.71 (0.51)	1.85 (0.74)	$.006^{**}$	.589	.892

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

3.4. Open-ended question responses

Table 4 summarizes the reasons participants in the two HMD groups gave for favoring either HMD-supported immersive learning or conventional learning. The most frequently mentioned reasons for preferring HMD was its immersive practice opportunities (10 respondents). For example, “I prefer the HMD environment because it creates an immersive experience without a real lab and creates motivation for learning” (Respondent #4). Many participants also mentioned ease of learning as a major reason for preferring HMD (five respondents). For example, “this way simulates real experiments to a great extent, and avoids the equipment problems of real electrical circuit operations. The experimental result is clearer and it’s easier to learn” (Respondent #14). Other reasons mentioned were the better learning effects (three respondents) and higher learning efficiency (three respondents) when using HMD. For example, “I can easily identify my learning issues and figure out the solution, which helps me to memorize and understand knowledge in a deeper way” (Respondent #7). The main reason given for preferring a conventional learning approach was that practice in a laboratory feels more real (four respondents). For example, “I like the lab environment, because I can learn how to operate in a real situation and learn the principles of electrical circuits more clearly” (Respondent # 40). Overall, HMD was more popular than conventional learning.

Table 5 presents the pros and cons of HMD-supported immersive learning according to the participants in the two HMD groups. Nine positive and seven negative aspects of HMD emerged in the feedback. The most frequently mentioned advantage of HMD was its authentic learning contexts (eight respondents), followed by its overcoming the constraints imposed by access to laboratories and materials (seven respondents), ease of learning (seven respondents), and the fun of learning through HMD (five respondents). For example, HMD "simulates the operation of a real experiment and allows experiential learning of knowledge" (Respondent #14); "convenient; we can have many choices, while lowering the cost in terms of time, space, and money" (Respondent #16); and "very interesting!" (Respondent #3).

Table 4. Main reasons for preferring either HMD-supported immersive learning or conventional learning

Immersive learning	Freq.	Conventional learning	Freq.
Immersive practice opportunities	10	More real	4
Easy to learn	5	Better for learning conceptual knowledge	2
Better learning effects	3	More convenient for operation	2
Higher learning efficiency	3	More familiar	1
Safer way of learning	2		
More interest in learning and higher motivation	2		
More engaged in learning	1		
Saves resources	1		

However, the participants were also concerned about eye discomfort (seven respondents), a less real experience than laboratory practice (six respondents), unstable positioning tracking (five respondents), and heavy equipment (five respondents). For example, "sometimes I feel uncomfortable when using it for a long period of time" (Respondent #6); "but still there is a sense that it is not real, because sometimes I find the image is unstable or vibrating when moving or staying still" (Respondent #11); "we do not need to go to the lab, which is more convenient, but the experience of HMD is not as good, and our vision becomes blurred and it requires more effort to learn" (Respondent #38); and "manipulation using handheld controller is not real enough, and the head-mounted display is too heavy" (Respondent #16).

Table 5. Main reasons for preferring either HMD-supported immersive learning or conventional learning

Pros	Freq.	Cons	Freq.
Experience authentic learning contexts	8	Eye discomfort	7
Overcome the constraints of access to labs and materials	7	Less real than lab practice	6
Easy to learn	7	Unstable positioning tracking	5
Fun	5	HMD is heavy and not very user-friendly	5
Practice-based learning	3	Not a wireless device and not safe	3
Flexible to operate	3	High cost and not easy to scale up	2
Better learning outcomes	2	Not real and may cause VR addiction	2
Just-in-time feedback	1		

4. Discussion

4.1. Effects of HMD and planning on problem-solving performance

The main findings partially confirmed our hypotheses that HMD and planning interventions help students to perform better in electrical-circuit design tasks and to transfer the knowledge to real-world settings. The HDM groups had higher success frequency and performed more efficiently than the 2D simulation groups in all four problems. Meanwhile, the groups that engaged in planning had better performances than groups that did not plan. Even for problems that showed no significant differences between groups, the descriptive statistics suggested a general trend in favor of HMD and planning interventions. Overall, the results supported the effectiveness of HMD and planning strategies for learning electrical-circuit design. The findings suggest that well-structured HMD-based simulation learning is a promising approach for use in various practical fields of engineering education.

Moreover, planning strategy having a larger effect than immersive media corroborated the argument that complex problem solving demands high regulative capacities of learners, which is one the paramount concerns in discovery learning (de Jong et al., 1998; Unterrainer & Owen, 2006). Whereas, immersive VR extends traditional discovery learning through offering high fidelity representation and a more rich and natural way of interaction, which supports knowledge understanding and knowledge construction but is better to be integrated with theoretically sound instructional strategies (Meyer et al., 2019). Therefore, our study reinforced Fowler's

(2015) position that immersive learning experience can be embedded within face-to-face instruction or real lab practice through rigorous learning design.

4.2. Effects of HMD and planning on the transfer of learning

It is especially intriguing to find that the HMD groups performed more efficiently than the 2D simulation groups in the post-transfer test involving real-world tasks, although there were no significant differences in the performances in the simulated training environments. This is consistent with previous studies showing that the user control afforded by HMD environment results in better transfer results (Gegenfurtner, Quesada-Pallarès, & Knogler, 2014), perhaps because HMD-based simulation environments allow users to naturally move their bodies, for example to walk, move their heads, grab or navigate some objects, which allows them to positively control the information received during the experience which, in turn, strengthens embodied learning (Hsu, Tseng, & Kang, 2018).

However, unlike a previous study (Meyer et al., 2019), we found no interaction effect between media and instructional design, perhaps because the participants in the non-planning groups performed some implicit planning even without a purposeful thinking-aloud process. Another possible explanation is that participants in the HMD groups had a heavier mental workload than those in 2D simulation groups, and planning made had no significant effect on reducing their workload. Further studies are needed to determine which type of workload, intrinsic, external, or germane load, occurs in simulation environments and under what conditions, so as to provide more effective contingent scaffolding in HMD-based learning environments (Makransky et al., 2019b). We elaborate on the workload issue in HMD in the next section.

4.3. Perceptions of HMD in support of problem solving

The students' opinions of the usefulness of HMD for learning were consistent with those identified in previous studies (Concannon et al., 2019; Makransky et al., 2019a) including both positive aspects, such as sense of presence, self-efficacy, simulator acceptance, and negative aspects, such as simulator sickness and mental workload. Their feedback on present feeling, simulator acceptance, and mental workload suggested that the visual representations in HMD and WIMP (window, icon, menu and pointer) in 2D simulation may have different effects on the learning experience (Barricelli, Gadia, Rizzi, & Marini, 2016; Jin & Lee, 2019). Further, according to participants' feedback, HMD was especially suitable for cultivating interest in learning and provided adequate opportunities for practice with just-in-time feedback, which is critical for novice learners to establish problem-solving confidence. Regarding the benefits of cultivating learning interest, supporting learning engagement, and improving self-efficacy in introductory electrical circuit design, the novel approach seems ideal for pre-service STEM teacher training for the learning experience may influence their future teaching practice.

However, they also expressed major concerns with HMD's technical flaws, such as eye discomfort, heavy equipment, and imperfect positioning tracking stability. Although HMD seems to be a better simulation environment than desktop 2D, some learners preferred to practice in a real laboratory. The divergent learning perceptions of using HMD implies whether to adopt the HMD-based problem-solving learning is not simply a yes/no question. This field deserves future research with respect to cost-effectiveness analysis regarding the adoption of HMD in electrical circuit design, as well as exploring innovative approach of integrating desktop simulation, HMD-based simulation with real lab experiments to optimize the benefits from different learning technologies.

The obvious perception of a heavy mental workload in HMD groups can be explained by the challenge of integrating conceptual knowledge, procedural knowledge, and psychomotor skills. For example, if learners received feedback from the HMD environment indicating a problem with their circuit solution, they had to scrutinize the circuit and reflect on their conceptual knowledge in the circuit design, generate troubleshooting strategies to locate the malfunctioning part of the circuit, orient the breadboard, and examine the connectivity of the electrical components to rule out the possibility of an open/short circuit. Such a reality-based interaction process demands a much higher comprehensive competence than 2D simulation. However, as Jost et al. (2019) have argued, this can be beneficial for building an intrinsic cognitive load in electrical engineering education, for the training goal is to perfect the cognitive processes and psychomotor skills in electrical design, wiring, testing, and troubleshooting.

5. Conclusions

This study investigated the effect of HMD and planning on problem-solving-based learning. In general, students who learned through HMD performed better than those who learned through 2D simulation, and students who adopted a planning strategy performed better than those who did not, although in some cases, the results failed to reach statistical significance. In addition, HMD-based learning induced a higher sense of presence, self-efficacy, and simulator acceptance than 2D simulation learning. However, the former media intervention also led to a heavier mental workload and more simulator sickness than the latter. In addition, although physical ergonomics of HMDs has improved a lot since its inception, technical problems related to wearing comfort, visual quality, positioning tracking and natural interaction were still a major concern for the learners. These findings imply that HMD is not a panacea that can support all levels of learning or that can replace laboratory-based problem-solving practice; however, it is a promising complement to the traditional way of learning disciplinary problem solving (Jaakkola et al., 2011). Besides, the study suggests HMD can be useful learning technology for similar disciplines like electrical engineering that requires hands-on practice to help students align domain knowledge with psychomotor skills. But it is also necessary to further address the side-effects of HMD-based learning, such as mental workload and simulator sickness, and to develop effective instructional strategies in immersive learning contexts.

This study has several limitations. First, it should be pointed out that the small sample size of this study limits the strength and generalization of the results; nonetheless, it does give a first glimpse on understanding the impacts of HMDs in problem-solving-based learning. Second, it would be worth investigating the problem-solving behavioral patterns of using head-mounted displays, such as eye movement, to increase our knowledge of mental workload in immersive virtual reality. Third, although 2D simulators are common in this field, and most have similar functionality and have been widely adopted in electrical engineering education, we only found one comparable HMD application in this field, and the functionality of these two simulation tools might not be exactly the same, such as types of electrical elements and circuit simulation feedbacks. To make sure the two simulation tools can provide the same constraints and functions, the problem-solving tasks were carefully selected and tested in both simulation conditions by a subject teacher. Finally, the long-term effects of HMD on the development of problem-solving competence needs to be investigated via a longitudinal study in the future.

Acknowledgement

This work is supported by the "Flower of Happiness" Pilot Research Fund (No. 2019ECNU-XFZH014), the Peak Discipline Construction Project of Education from East China Normal University, and the National Natural Science Foundation of China (No. 61977023). The authors would thank Jing Nie, Jiahui Ruan, Xinyu Chang, and Tianqi Zhang for their support of experiment implementation for this study.

References

Barricelli, B. R., Gadia, D., Rizzi, A., & Marini, D. L. R. (2016). Semiotics of virtual reality as a communication process. *Behaviour & Information Technology, 35*(11), 879–896.

Bauwens, S., & Ho, C. (2018). *Short circuit VR.* Retrieved from https://shortcircuitvr.com/

Chang, C. W., Heo, J., Yeh, S. C., Han, H. Y., & Li, M. T. (2018). The Effects of immersion and interactivity on college students' acceptance of a novel VR-supported educational technology for mental rotation. *IEEE Access, 6*, 66590-66599.

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics—Simulation and Computation, 39*(4), 860–864.

Chen, Y.-L., Hong, Y.-R., Sung, Y.-T., & Chang, K.-E. (2011). Efficacy of simulation-based learning of electronics using visualization and manipulation. *Educational Technology & Society, 14*(2), 269–277.

Concannon, B. J., Esmail, S., & Roberts, M. R. (2019). Head-mounted display virtual reality in post-secondary education and skill training: A Systematic review. *Frontiers in Education, 4*, 1–23.

Cooper, G., Park, H., Nasr, Z., Thong, L. P., & Johnson, R. (2019). Using virtual reality in the classroom: Preservice teachers' perceptions of its use as a teaching and learning tool. *Educational Media International, 56*(1), 1-13.

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

Falloon, G. (2020). From simulations to real: Investigating young students' learning and transfer from simulations to real tasks. *British Journal of Educational Technology, 51*(3), 778-797.

Field, A., Miles, J., & Field, Z. (2013). *Discovering statistics using R*. London, UK: Sage Publications.

Fowler, C. (2015). Virtual reality and learning: Where is the pedagogy? *British Journal of Educational Technology, 46*(2), 412-422.

Gegenfurtner, A., Quesada-Pallarès, C., & Knogler, M. (2014). Digital simulation-based training: A Meta-analysis. *British Journal of Educational Technology, 45*(6), 1097–1114.

Hart, S., & Staveland, L. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Advances in Psychology, 52*, 139–183.

Howell, D. C. (2006). *Statistical methods for psychology* (6th ed.). Belmont, CA: Duxbury.

Hsu, W. C., Tseng, C. M., & Kang, S. C. (2018). Using exaggerated feedback in a virtual reality environment to enhance behavior intention of water-conservation. *Educational Technology & Society, 21*(4), 187–203.

Huang, C. L., Luo, Y. F., Yang, S. C., Lu, C. M., & Chen, A. S. (2020). Influence of students' learning style, sense of presence, and cognitive load on learning outcomes in an immersive virtual reality learning environment. *Journal of Educational Computin*g Research*, 58*(3), 596-615.

Jaakkola, T., Nurmi, S., & Veermans, K. (2011). A Comparison of students' conceptual understanding of electric circuits in simulation only and simulation-laboratory contexts. *Journal of Research in Science Teaching, 48*(1), 71–93.

Jin, Y., & Lee, S. (2019). Designing in virtual reality: A Comparison of problem-solving styles between desktop and VR environments. *Digital Creativity, 30*(2), 107–126.

Johri, A., & Olds, B. M. (2011). Situated engineering learning: Bridging engineering education research and the learning sciences. *Journal of Engineering Education, 100*(1), 151–185.

Jost, P., Cobb, S., & Hämmerle, I. (2019). Reality-based interaction affecting mental workload in virtual reality mental arithmetic training. *Behaviour & Information Technology*. Advance online publication. doi:10.1080/0144929X.2019.1641228

Jung, J. K., & Ahn, Y. J. (2018). Effects of interface on procedural skill transfer in virtual training: Lifeboat launching operation study. *Computer Animation and Virtual Worlds, 29*(3-4), e1812. doi:10.1002/cav.1812

Kani, N. H. A., & Shahrill, M. (2015). Applying the thinking aloud pair problem solving strategy in mathematics lessons. *Asian Journal of Management Sciences and Education, 4*(2), 20–28.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology, 3*(3), 203–220.

Kwon, C. (2019). Verification of the possibility and effectiveness of experiential learning using HMD-based immersive VR technologies. *Virtual Reality, 23*(1), 101-118.

Lamb, R., Antonenko, P., Etopio, E., & Seccia, A. (2018). Comparison of virtual reality and hands on activities in science education via functional near infrared spectroscopy. *Computers & Education, 124*, 14-26.

Makransky, G., Borre-Gude, S., & Mayer, R. E. (2019a). Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning, 35*, 691–707.

Makransky, G., & Lilleholt, L. (2018). A Structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research and Development, 66*(5), 1141-1164.

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019b). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction, 60*, 225-236.

Mayer, R. E., & Wittrock, M. C. (2006). Problem solving. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (2nd ed.) (pp. 287–304). Mahwah, NJ: Erlbaum.

Meluso, A., Zheng, M., Spires, H. A., & Lester, J. (2012). Enhancing 5th graders' science content knowledge and self-efficacy through game-based learning. *Computers & Education, 59*(2), 497–504.

Meyer, O. A., Omdahl, M. K., & Makransky, G. (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education, 140,* 103603. doi:10.1016/j.compedu.2019.103603

Miles, M. B., & Huberman, M. (1994). *Qualitative data analysis: A Sourcebook of new methods* (2nd ed.). Beverly Hills, CA: Sage Publications.

Mills, J. E., & Treagust, D. F. (2003). Engineering education—Is problem-based or project-based learning the answer. *Australasian Journal of Engineering Education, 3*(2), 2–16.

Nokes-Malach, T. J., & Richey, J. E. (2015). Knowledge transfer. In R. A. Scott & M. C. Buchmann (Eds), *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource* (pp. 1-15). Hoboken, NJ: John Wiley & Sons, Inc.

Nunnally, J. C. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.

Parmar, D., Bertrand, J., Babu, S. V., Madathil, K., Zelaya, M., Wang, T., Wagner, J., Gramopadhye, A. K., & Frady, K. (2016). A Comparative evaluation of viewing metaphors on psychophysical skills education in an interactive virtual environment. *Virtual Reality, 20*(3), 141-157.

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology, 110*(6), 785-797.

R Core Team. (2019). *A Language and environment for statistical computing*. Retrieved from https://www.r-project.org/

Richards, D., & Taylor, M. (2015). A Comparison of learning gains when using a 2D simulation tool versus a 3D virtual world: An Experiment to find the right representation involving the marginal value theorem. *Computers & Education, 86*, 157-171.

RStudio. (2019). *Integrated development environment (IDE) for R*. Retrieved from https://rstudio.com/

Shah, D. (2018). *Breadboard Simulator v1.0*. Retrieved from https://ds0.me/csim/index.html

Shu, Y., Huang, Y. Z., Chang, S. H., & Chen, M. Y. (2019). Do virtual reality head-mounted displays make a difference? A Comparison of presence and self-efficacy between head-mounted displays and desktop computer-facilitated virtual environments. *Virtual Reality, 23*(4), 437–446.

De Jong, T., Van Joolingen, W. R., Swaak, J., Veermans, K., Limbach, R., King, S., & Gureghian, D. (1998). Self-directed learning in simulation-based discovery environments. *Journal of Computer Assisted Learning, 14*(3), 235- 246.

Unterrainer, J. M., & Owen, A. M. (2006). Planning and problem solving: From neuropsychology to functional neuroimaging. *Journal of Physiology - Paris, 99*, 308–317.

Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences, 39*(2), 273–315.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A Presence questionnaire. *Presence, 7*, 225–240.

Wu, B., Yu, X., & Gu, X. (2020). Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. *British Journal of Educational Technology.* Advance online publication. doi:10.1111/bjet.13023

Wu, X. Q., Liu, T. L., & Yi, Z. A. (2012). Introduction research and practice of training mode reform in the higher engineering education. In L. Zhang & C. Zhang (Eds.), *Engineering Education and Management* (pp. 33-37). Berlin, Germany: Springer.

Yang, X. Z., Lin, L., Cheng, P. Y., Yang, X., Ren, Y. Q., & Huang, Y. M. (2018). Examining creativity through a virtual reality support system. *Educational Technology Research and Development, 66*(5), 1231-1254.

Zacharia, Z. C. (2007). Comparing and combining real and virtual experimentation: An Effort to enhance students' conceptual understanding of electric circuits. *Journal of Computer Assisted Learning, 23*(2), 120-132.

Zhang, H., Yu, L., Ji, M., Cui, Y., Liu, D., Li, Y., Liu, H., & Wang, Y. (2020). Investigating high school students' perceptions and presences under VR learning environment. *Interactive Learning Environments, 28*(5), 635-655. doi:10494820.2019.1709211

Appendix 1

Construct	Items	Source
Sense of presence	1. I could easily engage into this simulation environment. 2. I was very engaged in solving the circuit design problems and barely noticed anything around me. 3. I focused on completing the problem-solving tasks and did not realize the passage of time. 4. I had a feeling of being in a circuit lab.	(Witmer & Singer, 1998)
Self-efficacy	1. I believe I can complete the task of circuit design experiment. 2. I believe I have developed good knowledge and skills for completing circuit design experiments. 3. I believe I have mastered the knowledge and skills of circuit design experiments.	(Meluso et al., 2012)
Mental workload	1. I can operate freely when I conduct a circuit design experiment in the simulation environment. 2. I can easily complete the circuit design task in the simulation environment. 3. I don't need to remember or think too much when I work on the problems in the simulation environment. 4. I am more relaxed in the process of completing the experiment in the simulation environment.	(Hart & Staveland, 1988)
Simulator sickness	1. I often feel unwell throughout the circuit design task in the simulation environment. 2. I often feel dizzy throughout the circuit design task in the simulation environment. 3. I often feel blurred vision throughout the circuit design task in the simulation environment.	(Davis, 1989), (Venkatesh & Bala, 2008)
Simulator acceptance	1. The simulation environment can help me learn electrical circuit design. 2. The simulation environment makes it easier for me to understand problem-solving task as well as to learn task-related knowledge and skills. 3. I am satisfied with circuit design study in the simulation environment. 4. I am more interested in learning circuit design through the simulation environment. 5. I am willing to learn relevant knowledge and skills in such a simulation environment in the future.	(Kennedy et al., 1993)

Integrating Games, e-Books and AR Techniques to Support Project-based Science Learning

Yi Hsuan Wang
Department of Educational Technology, Tamkang University, Taiwan // annywang12345@hotmail.com

(Submitted April 22, 2020; Revised May 30, 2020; Accepted August 5, 2020)

ABSTRACT: The study integrated learning technology and various e-learning materials to assist teachers in conducting project-based learning for a science course. The activities were designed for learning science concepts such as circuits and the symbols of electricity firstly, and then for applying the concepts to produce a scientific toy. The study intended to integrate the value of the convenience and accessibility of e-books, the feature of interactive demonstration, a combination of visual information with physical objects using augmented reality (AR), and the enjoyment brought by game-based learning to create an integrated e-learning model to support project-based learning processes. A total of 51 elementary school students were invited to take part in the research and were separated into two groups, one using the game-based learning environment with AR-based materials and the other with e-books for the science project-based activity. Despite the quantitative data not presenting any learning differences for the two groups, the qualitative results showed that the e-learning materials with multimedia content were helpful for scaffolding students while completing the hands-on scientific electric current toy, and triggered peer discussion to achieve concept agreement. For example, the text and figures in the e-books helped the students to doubly confirm the processes of the learning information, while the colored blocks on the physical objects in the AR materials facilitated assembly of the elements to complete the science toy. It was noticed that the well-designed AR materials might hinder learners' thinking ability and restrict their creativity, and hence, the researcher proposes a revised e-learning integration model that combines the advantages of e-books and AR techniques to create more easily accessible e-learning supports and to encourage active thinking for fostering independent and self-regulated learners.

Keywords: Interdisciplinary projects, Project-based learning, Game-based learning, AR support learning

1. Introduction

Project-based learning emphasizes giving students chances to independently seek answers to questions and to solve problems via utilizing the learned knowledge. It also provides learners with opportunities to acquire manual skills through performing activities (Chen, 2004). Project-based learning has the potential to assist students in learning science (Panasan & Nuangchalerm, 2010) since a project-based science classroom encourages learners to explore phenomena, build up their own knowledge structure, try out new ideas and solve the problems they face (Remziye & Kargın, 2014). The key features of project-based activities are starting with a driving question and asking students to explore the question through inquiry processes. The students then learn and apply the idea to find solutions and are scaffolded by the learning technology during the inquiry processes (Krajcik & Shin, 2014). A good project-based activity would induce learners to progressively achieve higher order learning abilities at the levels of Bloom's taxonomy (Anderson & Krathwohl, 2001), where they use their basic level knowledge (remember and understand) to first solve the problems (apply and analyze), and then to produce concrete products (evaluate and create). However, it was noticed that when adopting project-based learning instruction in a course, teachers might face the challenge of limited course time, a tight course schedule, and heavy loading of course preparation (Lawson, 1995).

Researchers have argued that the abovementioned dilemma faced by teachers might be alleviated through integrating learning technology to assist instruction (Shapley et al., 2011). For example, e-books with interactive demonstration provide students with opportunities to interact with the learning content (Smeets & Bus, 2012), and using e-books with various learning methods such as the flipped classroom strategy can benefit learners' learning efficacy and achievement (Wang & Huang, 2017). The affordances of augmented reality (AR), a combination of visual information with physical objects, are helpful for presenting and explaining abstract scientific phenomena (Chen et al., 2018). The AR superimposing virtual elements onto real-world environments with visualized details provides scaffolds to help students acquire abstract science concepts (Yoon et al., 2017). Meanwhile, students have the freedom to observe phenomena, fail, experiment, and interpret in the game-based learning environment, allowing them to increase their physical skills (Klopfer et al., 2009), clarify their conceptual understanding (Barak & Dori, 2005), and promote their learning motivation (Rosen, 2009). Hence, games can serve as tools for scientific learning (Tsai & Tsai, 2020).

In short, the project-based learning approach has been shown to have positive effects on students' academic performance since it involves the transformation and construction of knowledge. It gives learners opportunities to transform and construct knowledge through asking questions, seeking answers, and performing investigations. Besides, in order to help students perform project-based activities effectively, the learning technology could be adopted in the classroom to lead more effective project-based instruction (Ozdamli & Turan, 2017), motivate learners' interest, and support them to achieve better academic performance (Chen & Yang, 2019).

1.1. Research purpose and questions

Understanding the above research background, this study aimed to conduct multi-phase research to propose a possible technology-enhanced integration model to support project-based science learning. The project-based activities in the study were designed and the technology-based learning materials were developed to support teachers with diverse curriculum materials for conducting science project-based learning activities. During the project-based learning processes, various technology-based learning materials were integrated according to their characteristics including (1)the convenience and accessibility of the e-books; (2)the nature of games that create an environment in which learners are free to explore knowledge in a low-stress and happy atmosphere, and (3)the characteristics of AR-based applications that combine virtual and real environments to offer students interactive and inquiry-based learning, adopted as scaffolding to help students acquire different cognitive knowledge levels of science concepts. The research questions of the study are:

- Are there differences in the learning performance of the students who learn through the game with e-book learning materials and the game with AR-based learning materials?
- What are the students' and teacher's feedback on using the game with e-book and the game with AR-based learning materials for learning?

2. Literature review

2.1. Project-based learning

The concept of project-based learning comes from Dewey's (1938) *learning by doing,* which is a student-centered model that organizes learning around projects or authentic learning experiences. Project-based learning encourages students to investigate questions, and to propose and explain the ideas through engaging in the activities as a form of situated learning that is based on constructivism. Constructivism encourages students to actively construct their own knowledge and reflect on how their understanding is changing (Elliott et al., 2000). The design of project-based learning is complex and is based on challenging (driving) questions or problems, for example, "How does science impact people's lives?" It involves students in learning by design, problem-solving, decision making, or investigative hands-on activities. The students could thus encounter and learn the discipline via projects (Marx et al., 1994). Krajcik and Shin (2014) summarized several key elements of project-based learning. First, start with a driving question. Use a good driving question to provide students with a context in science practice and let them build meaningful understandings of key scientific concepts when pursuing solutions to the driving question. Researchers have suggested that anchoring experiences present meaningful contexts for the science ideas explored in the project and show the value of the project's driving question. Second, encourage situated scientific inquiry and collaborations. Third, use technology tools to support learning. The learning technology tools enable learners to access information on the World Wide Web, extending to what could be done in the classroom. For example, the use of learning technology supports natural phenomenon simulation in an easier way within the limited amount of time. With the help of learning technology tools, it is possible for teachers to move away from the transmission and acquisition model of instruction, so that they are not the only ones to give knowledge, but the students can explore the environment to construct their own knowledge. Last, create a product. When students build artifacts, it enhances their understanding because they have to tie together science concepts to support the product. Besides, the teacher uses artifacts to know how students learn across various projects. Researchers have explored the effects of adopting project-based instruction in learning; for example, Remziye and Kargin (2014) made a comparison between using the project-based approach and a teacher's original method in the learning of the Electricity unit for elementary sixth graders. They found that the students succeeded more through project-based learning in the achievement tests because of their active participation.

The reviewed research has concluded that learners could promote their learning ability by proposing and investigating problems, and apply their knowledge in explaining and verifying ideas. Through these repeat processes of project-based activities, the learners would be able to progressively achieve higher order learning abilities from basic level knowledge to higher order learning skills, thus having positive effects on their learning.

2.2. Project-based science learning

Educators have noticed that learners might not be motivated to learn science since there is a lack of opportunities to help them develop their explanations of the real-world phenomena or because of the ineffective textbook design and instructional style that provide students with cookbook-like instructions, asking them to follow the procedures without having a deeper understanding of the materials (Krajcik & Shin, 2014). A project-based science classroom would be able to avoid the abovementioned situation since a project-based science course encourages learners to explore, investigate, discuss, implement and modify their ideas though trial and error procedures. According to Bloom's taxonomy, there are six levels of learning objectives in the cognitive domains of remember, understand, apply, analyze, evaluate and create (Bloom et al., 1956). A well-designed project-based science activity would induce learners to acquire knowledge from lower-order skills that require less cognitive processing to higher-order skills that require deeper and complex cognitive processing. Besides, project-based science learning would create contexts and driving questions to encourage learners to explore their idea, find solutions and have discussions to demonstrate how they understand and apply the learned knowledge (Marx, 1998). One study observed that the use of the project-based method increases learners' success in science (Ergül & Kargın, 2014), and found that students who did a project-based learning activity were able to produce new knowledge on their own investigation and exploration. Meanwhile, teachers and peers worked together and made conceptual advances through exchanging views (ChanLin, 2008). In the project-based learning scenario, students shared what they discovered during the co-operative activities and their disagreement aroused discussions that made them learn from each other (Sawyer, 2004).

It was noticed that the project-based course takes comparatively more time for teachers and students to design and complete the learning tasks since the processes of knowledge construction are not easy to perform. Moreover, instructors have revealed the difficulty of improving learners' motivation and making them concentrate on learning tasks, especially when project-based activities are applied in large classes (Sumarni, 2015). From this point of view, technology tools would be helpful because learning technologies allow students to access real data and extend what they can do in the classroom for actively constructing knowledge.

2.3. Technology-supported project-based science learning

The use of learning technology tools presents learning information in more dynamic and interactive formats that could serve as powerful instructional tools to help teachers more effectively foster inquiry learning (Nielsen et al., 2016). For example, some science phenomena, such as magnetic fields, atoms or galaxies, cannot be easily observed, and researchers have indicated that computerized visualization or animation to present and explain abstract scientific phenomena is more effective and permanent than traditional learning (Dori & Belcher, 2005). Besides, traditional hands-on experiments in science laboratories could be broadened or enhanced through integrated simulation and animation tools (Krajcik & Mun, 2014). It has been argued that teaching physics through computer-based simulations, 2D-based animations or animated movies makes the content more easily graspable (Barak et al., 2011;).

Several studies have integrated the learning technology into support project-based learning activities, and have suggested that technology-supported learning could be a good way for teachers to create more interactive and student-centered learning when integrating project-based activities into courses (Ozdamli & Turan, 2017). For example, Eskrootchi and Oskrochi (2010) found that using a computer-based project-based environment for simulation helped to improve students' understanding of the science watershed concept learning. Ozdamli and Turan (2017) adopted a blended learning environment to support a project-based science learning activity for college students taking a mobile application development course, and explored the effects on students who learned with and without the technology supported project-based learning steps. They found that these students showed better learning success than those learning with traditional methods. Sung and Wu (2018) explored the effects of integrating project-based activities with or without e-books to support a nursing course for learning health concepts, and found that the comprehension and ability of the learners with the e-books were improved, and learners' cognitive skills and problem-solving ability were improved via the project-based learning. Using technology tools during project-based learning enables learners to access information easily on the World Wide Web, extending what could be done in the classroom (Krajcik & Shin, 2014), and creating more interactive and student-centered learning (Chauhan, 2017).

2.4. The learning media: games, e-Books and AR for learning

Learning media have evolved quickly with the development of learning technologies. The value of the convenience of e-books, a combination of visual information with physical objects using AR, and the enjoyment brought by game-based learning are new media which in some sense subsume earlier media (Collins, 1994).

E-books, also known as e-textbooks, digital-textbooks or electronic textbooks, embed multimedia and have more interactive functions than traditional textbooks (Weng et al., 2018). The value of the accessibility of the e-books makes them welcome by most teachers and students (Embong et al., 2012). Meanwhile, the creation of individualization and learning based on students' self-paced learning, and the embedding of an appropriate number of multimedia in e-books can help learners to improve their learning (Spanovic, 2010). Weng et al. (2018) stated that good utilization of e-books provides students with chances to engage collaboratively. Hwang and Lai (2017) adopted e-books to support flipped learning, and revealed that the use of e-books with interactive learning content facilitates in- and out-of-class learning. Chen and Su (2019) used an e-book reading system to trace students' learning progress, and found that the system supported students' self-regulated learning and self-efficacy.

Game-based learning is a highly exciting medium which refers to using game elements to engage students in tasks for achieving learning goals (Liao et al., 2013). The game-based learning environment provides students with a wide range of ways to experience trial-and-error processes, and they could hence achieve self-regulated learning in solving the tasks during gameplay (Plass et al., 2015). Meanwhile, learning occurs during gameplay providing a powerful affordance for supporting learning interaction, increasing learning motivation and promoting learning performance (Gee, 2003). Researchers (Liao et al., 2013) have revealed that students improved their science concepts and performance with the game-based practice, and that the game-based interaction was effective in terms of engaging students in discovering environments and enacting problem solving.

The AR-based learning materials merge virtual information with real objects and present virtual information such as text, video, audio, and three-dimensional learning content for instruction interaction. The characteristics of AR-based application include the integration of static and dynamic content, and the combination of virtual and real environments plays a beneficial role in science education regarding concept change, inquiry-based learning, practical skills and scientific argumentation (Chen, Huang & Chou, 2017). Teachers have suggested that AR learning materials would be more helpful if they could be more flexible and controllable in terms of adding or removing elements to or from the AR content (Kerawalla et al., 2006), and it has been reported that AR technology offers promise for transforming science learning. For example, Cheng and Tsai (2013) explored the effects of integrating AR into supporting science learning, and claimed that image-based AR would benefit learners' practical skills and conceptual understanding, while location-based AR is more helpful for inquiry-based scientific activities. Matcha and Rambli (2013) developed an image-based AR application, AR Circuit, to assist students in investigating the relationship of the elements in an electric circuit, and found that the physical AR objects promoted learners' discussion and collaboration during science concept learning. Chen et al. (2017) presented a blended learning environment through AR techniques to help elementary school students understand the growth pattern of leaves in a science course, and concluded that AR-based blended learning was helpful for outdoor exploration activities. Yoon et al. (2017) found that learners using the AR technique demonstrated better knowledge gain on science because AR affords greater ability to visualize details and hidden information. Studies have indicated that using learning technology sustains learners' motivation and learning engagement (ChanLin, 2008) and leads to effective learning (Chauhan, 2017). Despite the fact that many studies have revealed the positive learning effects of using AR techniques for learning, some have argued that learners with AR supports were less engaged than those using traditional learning resources because they were asked to watch an AR animation and describe it passively instead of having the chance to do the exploring (Kerawalla et al., 2006).

2.5. Summary of the literature review

After reviewing the related literature, the possible advantages of project-based learning and the features of technology-supported learning have been presented. The project-based approach enables learners to learn multiple disciplines and progressively build their own learning from basic level knowledge to higher order learning skills during the completion of the tasks. Besides, supporting teachers with diverse curriculum materials for conducting science project-based learning activities is also important. The technology tools would be helpful to support project-based instruction, and the use of game-based learning can arouse students' learning motivation, and the adoption of e-books and AR-based content creates more interactive environments for self-exploring activities. Hence, in this study, the e-learning materials were developed with the aim of combining the features of technologies to support teachers in designing the project-based science activities and to scaffold students in acquiring science knowledge from the basic cognitive process dimension to a higher order level.

3. E-learning materials design and development

Better design of learning educational applications could be achieved by having teachers and educational technology experts work together in a collaborative process of development to reduce the gap between system designers and practitioner teachers. Hence, science instructors were invited to participate in the co-design stage for structuring the project-based activities and learning materials according to various learning purposes. The co-design stage lasted for about half a year in which the research data from interviews from the science teacher and class observation were collected as a pilot study to understand the teachers' needs and students' learning problems (Figure 1).

Figure 1. The process of content development

After analyzing the data from the co-design stage, it was found that the students' difficulties might include the concept of electric circuits or the characteristics of electricity such as comprehending or designing the real shapes of electric circuit diagrams. Hence, in order to assist students in adopting the learnt concepts into practice, the project-based activity and e-learning materials were designed and developed to help them acquire the concepts of basic electricity and then to implement the science concepts to produce a scientific electric current toy.

3.1. The structure of the materials

The learning content was designed for acquiring the basic electricity concepts such as electric circuits, the concepts of paths, open and short electric circuits and series and parallel, and the advanced knowledge such as combining the electrical elements and completing the scientific electric current toy. The learning content included two topics (Figure 2). Following Bloom's taxonomy, Topic1 was designed for acquiring the basic electricity concepts on the scales of the remember and understand knowledge dimensions, and Topic 2 was an advanced practical hands-on activity that asked students to apply learned science concepts in producing a scientific electric current toy for acquiring the scales of apply, analyze knowledge dimension. The game-based environment was created for Topic 1, basic concept learning, to provide learners with trial-and-error opportunities and to encourage them to do the learning inquiry through the game-based interaction. The e-books and AR-based interaction were created for Topic 2, the advanced practical hands-on activity, to give the students the necessary scaffolding to apply the previous basic science electricity concepts to complete the higher level cognitive learning tasks.

Topic 1: Basic concept learning (Remember and understand knowledge dimension)	**Topic 2: Advanced practical hands-on activity** (Apply, analyze, evaluate and create knowledge dimension)
1. To know the basic requirements of electric circuits 2. To recognize the concepts of path, open and short electric circuits 3. To know the concept of series and parallel 4. To recognize circuit diagrams, current, voltage, and resistance	1. To recognize the elements of buzzer, wire and breadboard 2. To combine the electrical elements and complete the scientific electric current toy

Figure 2. The topics and corresponding learning content

3.2. Topic 1

The purpose of Topic 1 was to acquire the basic science concepts of electrical circuits, series and parallel. There were three levels in the game. The first two were designed as multiple-choice interactions. The students explored the game-based environment with their own avatar. They had to follow the instructions and answer the questions correctly to overcome the challenges. The third level was a game-based examination, and the students had to use drag-and-drop interaction to fill in the blanks to complete the electricity diagram (Figure 3).

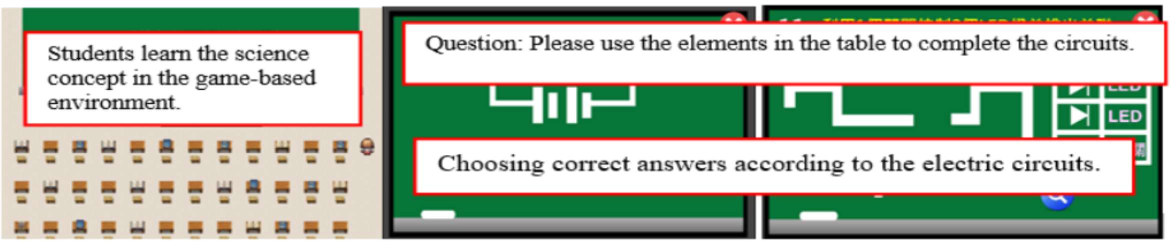
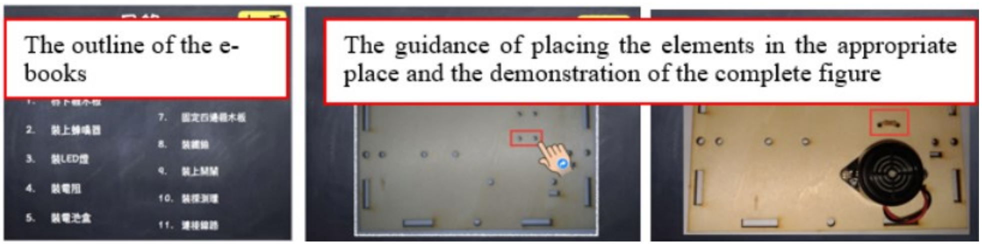


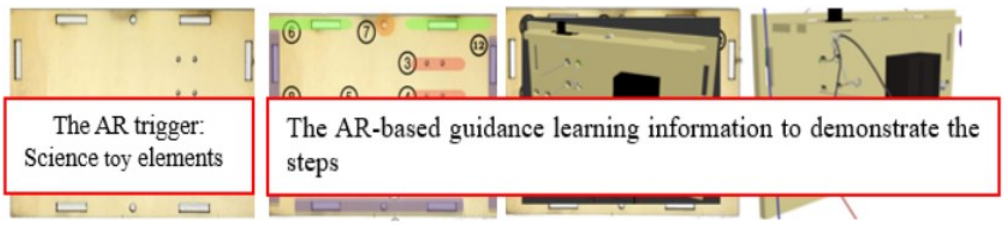
Figure 3. The game-based learning marerials

3.3. Topic 2

The purpose of Topic 2 was to encourage inquiry and to create a science toy, also called an electric current avoider. After the basic concepts were acquired, the students had to apply the previous basic science electricity concepts to complete the higher level cognitive learning tasks. The requirement of the science toy is that when the toy was powered on, the learners moved the wire ring along the other wire, and when the wire ring collided with another wire (indicating that the path was connected), the buzzer would sound. The e-books and AR-based materials were used as scaffolding to provide the students with guiding steps when they produced the science toy (Figure 4). The content and guiding steps in the e-books and AR-based content were the same. The e-books provided the learners with figures and textual instruction, while the AR-based content provided them with more interactive instruction. The students could use the learning devices to scan the science elements, and then the instruction including textual guidance, video, and 3D animation would show up on the screen.



(a) The e-book learning content



(b) The AR-based learning content

Figure 4. Print-screens of the learning materials

4. Methodology

4.1. Participants

A total of 51 elementary school students participated in the study and were randomly separated into two groups. A general science concept test was conducted to confirm that the learners of the two groups had an equal learning starting point. Group A (GA) consisted of 30 students who learned with the game-based learning materials and e-books, and Group B (GB) consisted of the other 21 students who learned with the game-based learning materials and AR.

4.2. Research design and procedure

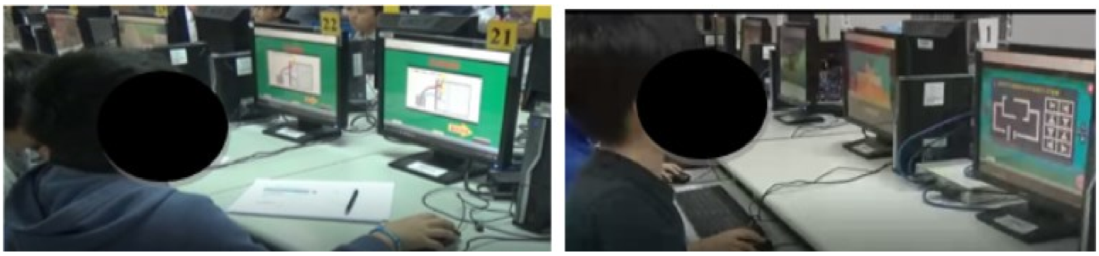
The duration of the experiment was 4 weeks, at 120 minutes per week, and the two groups were taught by the same science teacher (Figure 5). The Topic 1 activity took 2 weeks, and the students in the two groups learned the basic electricity concepts from the teacher's instruction. Both groups were able to practice with the game-based materials (Figure 6a) to acquire lower level knowledge (remember and understand).

Phase 2: Course Experiment

Weeks	Topic	Group A	Group B
1	Topic 1	Game-based learning materials	
2			
3	Topic 2	E-books	AR-based materials
4			

Figure 5. The processes and e-learning materials arrangement for the group in the experiment

Then, the Topic 2 activity took another 2 weeks in which the learners had to apply the learned knowledge to produce the scientific toy to achieve higher level knowledge (apply and analyze). During the Topic 2 hands-on session, the learners produced their science toy in the science classroom where there were several long rectangular tables, each occupied by four to five students. The teacher first demonstrated the complete version of the science toy to the students and briefly reviewed the basic concepts used in designing the toy. Then, each student got a material package including the components of boards, wires, buzzers, batteries and screws, and they had to make the science toy on their own. The students in GA were given e-books (Figure 6b) as their learning supports, while the GB students used the AR-based materials as supports (Figure 6c).



(a) Topic 1: game-based learning materials

(b) Topic 2:The e-books (c) The AR-based learning materials

Figure 6. The students doing hands-on activities with the e-learning materials

4.3. Data collection and analysis

The collected research data included students' pre- and post-tests, their hands-on activity scores, questionnaires and interview data.

4.3.1. The pre- and post-tests

The pre- and post-tests consisted of multiple-choice questions for testing the learners' basic science concepts and an open-ended question for testing their understanding of circuit connections. The test questions were designed from the school's textbook organization and course materials, and the science teacher had reviewed and confirmed that the wording and the level of the items fulfilled the learning purpose. The test was a summative evaluation that helped the researcher to examine how much the students had gained through the experiment. The total scores of the pre- and post-tests were each 100. In the multiple-choice questions, the students had to choose answers from four items to demonstrate their learning acquisition of the basic science concepts. In the open-ended question, the students had to write down and distinguish the elements of an electrical circuit.

4.3.2. The hands-on activity scores

Each student also had a hands-on activity score according to the level of completion of the scientific toy. The evaluation of the hands-on activity was graded by the science instructor according to five criteria, namely installation of the Switch, LED, Buzzer, the correctness of the circuit diagram and their debugging analysis, and each criterion was graded individually as the quantitative results according to three levels: 3 points (Full completion), 2 points (Half completion), and 1 point (No progress). The total score of the hand-on activities was 15 points.

4.3.3. The questionnaires

The research questionnaires were administered and an interview was carried out after the experiment. The questions in the questionnaires included various subscales with items on a 5-point Likert scale (from 5 to 1: *strongly agree, agree, neutral, disagree, strongly disagree*) and open-ended questions to investigate how the learners perceived using the various learning materials for science project-based learning. The purpose of conducting the questionnaires was to understand how the students perceived using various technology-enhanced learning materials for science learning in terms of functions and interface design, and general feedback including learning stratification and motivation. The question items of the two groups were the same, but one more question was added for GB (the AR group) to understand the learners' feedback on the 3D object design in the materials. For the open-ended questions, the students were encouraged to write down their feedback on using the game, e-books and AR-based materials during the activity. The coefficient α of the measures of the questionnaires from GA and GB were .98 and .95, respectively.

4.3.4. The interview

The teacher and several students (5-7of each group) were invited to take part in an interview after the experiment so as to understand their perceptions and opinions of using the materials for learning. The teacher helped the researcher to pick the students for the interviews based on their performance during the course. Half of them were those who performed especially well and the others were randomly selected by the teacher. The interview questions for the teacher were: (1) How do you perceive integrating the various e-learning materials for the course? (2) After integrating the learning materials into the hands-on activity, do you have any preference for the e-books or the AR-based learning materials? Which might help the students better? The interview questions for the students were: (1) How did you feel when you practiced the science concepts with the game-based learning materials? (2) What did you do when you started to make the scientific toy? Did you follow the instructions in the materials step-by-step or did you try on your own first? Please share your experience. (3) Did the learning materials help you to conduct the science activity? Why or why not? (4) Do you have any suggestions regarding the design of the learning materials?

5. Results

The data were analyzed and are presented according to the research questions, and the quantitative and qualitative data, including the students' pre- and post-tests, hands-on activity scores, questionnaire and interview answers, were analyzed for methodological triangulation. Descriptive statistics were calculated to describe the means and standard deviations, and an independent samples t test and paired samples t test were adopted to compare the learning and questionnaire results between and within the two groups after the experiment. For the qualitative data from the instructor and students, each participant was given a code, group-sex-number, for example, GA-1-3 represents the data from learner 3 who is a boy in GA who learned with e-books and the game. GB-2-1 represents the data from learner 1 who is a girl in GB who learned with AR-based materials and the game. The researcher translated the feedback from the interviews into raw data files for each participant, and re-coded the raw data according to different themes. The final qualitative data were organized and displayed as reduced data from which the findings for each question could be highlighted.

5.1. Learning performance

For answering research question 1 (Are there differences in the learning performance of the students who learn through the game with e-book learning materials and the game with AR-based learning materials?), the scores of the pre- and post-test were used for the paired samples t test in order to understand how the students improved before and after the experiment. The scores of the post-test and hands-on activity were adopted for the independent samples t test to examine whether there were differences in the learning performance of the students from GA and GB. According to the results of the paired samples t test (Table 1), it was found that the GA and GB learners all improved their learning after the course, and the statistical results achieved significant difference (GA t = -11.088, p = .00; GB t = -5.445, p = .00) which indicated that both types of e-learning materials helped their learning. However, the learning performance of the two groups did not show any difference according to the results of the independent samples t test (Table 2, t = 1.218, p = .229). When further analyzing the learners' hands-on activity scores, it was found that the total scores of the two groups were quite close, and the statistical results of the independent samples t test did not show a statistical difference (Table 3, t = 1.848, p = .071). However, it was noted that when analyzing the hands-on activity in more detail, the learners in GA with the e-books performed better on the sub-item of installation of the Switch (Table 3, t = 2.230, p = .030) and the correctness of the circuit (Table 3, t = 2.286, p = .61) than the learners in GB, and the scores achieved a significant difference.

Table 1. Paired samples t test of the pre- and post-tests

Group A (GA)	Mean	*SD*	*t*	*p*
Pre-test	16.33	15.34	-11.08	$.00^{***}$
Post-test	62.67	16.91		
Group B (GB)	Mean	*SD*	*t*	*p*
Pre-test	24.19	19.89	-5.44	$.00^{***}$
Post-test	56.19	21.00		

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

Table 2. Independent samples t test of the pre- and post-tests

		Mean	*SD*	*F*	*t*	*p*
Post-test	Group A (GA)	62.67	16.91	4.46	1.21	.22
	Group B (GB)	56.19	21.00			

Table 3. Independent samples t test of the hands-on activity

		Mean	*SD*	*F*	*t*	*p*
Total score	GA	11.67	3.14	1.35	1.84	.07
	GB	9.90	3.63			
Sub-item	GA	2.60	0.72	2.68	2.23	$.03^{*}$
Installation of Switch	GB	2.10	0.88			
Sub-item	GA	2.27	0.90	0.41	1.03	.30
Installation of LED	GB	2.00	0.89			
Sub item	GA	2.60	0.77	0.10	1.65	.10
Installation of Buzzer	GB	2.24	0.76			
Sub-item	GA	2.20	0.76	0.84	2.38	$.02^{*}$
Correctness of circuit	GB	1.71	0.64			
Sub-item	GA	2.00	0.91	1.55	.581	.56
Debugging analysis	GB	1.86	0.79			

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

5.2. Quantitative results from the questionnaire

The data from the questionnaires were analyzed to answer the second research question (What are the students' and teacher's feedback on using the game, e-books and AR-based learning materials for learning?). Descriptive statistics and an independent samples t test were conducted to analyze the two groups' feedback after the experiment. The questionnaire results are presented in Table 4. The average scores of the items revealed that the students of the two groups had positive feedback on learning with the assistance of the e-learning materials, and most scores were above 4 points (on a 5-point Likert scale from 0 to 5). It was noticed that the learners with the AR-based learning materials had higher average scores for most questions than the scores of the learners with e-books (Table 4, Q1-Q4, and Q6-Q10) indicating that the students liked to use the AR-based learning materials and reflected that the AR-based learning supports helped them understand the learning concepts and complete the hands-on work independently. GB also revealed that the instruction from the 3D demonstration helped them complete the hands-on activity (Table 4, GB-Q11). It was also noticed that the average score of Question5 was higher for GA (Table 4, Q5, GA *Avg.* = 4.5, GB *Avg.* = 4.19), showing that the learners with the e-books as technology-enhanced learning materials agreed more with the help provided by the learning technology for the hands-on activity.

Table 4. The results of questionnaires: Part one

Part one questions	Group	*Avg.*	*SD*	*t*	*p*
1. The learning material was easy to use.	GA	3.83	1.31	-.81	.41
	GB	4.10	0.76		
2. The design of the interface was clear.	GA	4.33	1.18	-.16	.87
	GB	4.38	0.80		
3. The instruction of the material was clear.	GA	4.10	1.34	-.27	.78
	GB	4.19	0.75		
4. The use of the materials helped me to understand the learning concepts.	GA	4.20	1.15	-1.34	.18
	GB	4.57	0.59		
5. The use of the materials helped me to know how to use the science tools for the hands-on activity.	GA	4.50	1.07	1.02	.30
	GB	4.19	1.03		
6. With the help of the materials, I could complete the hands-on activity without the teacher's assistance.	GA	4.07	1.28	-1.33	.18
	GB	4.48	0.68		
7. The use of the materials promoted my motivation for conducting the hands-on activity.	GA	4.10	1.26	-.91	.36
	GB	4.38	0.74		
8. The use of the materials promoted the effectiveness of the hands-on activity.	GA	4.13	1.27	-.03	.97
	GB	4.14	0.65		
9.With the help of the learning materials, the hands-on activity was not so hard.	GA	3.80	1.42	-.60	.55
	GB	4.00	0.63		
10. I like to use the learning materials for learning	GA	4.13	1.25	-.79	.42
	GB	4.38	0.80		
*GB 18. The instruction from the 3D demonstration helped me to complete the hands-on activity.	GB(only)	4.24	0.88		

5.3. Qualitative results from the questionnaires and interviews

The feedback from the open-ended questions and interviews are organized in Table 5 to answer the second research question (What are the students' and teacher's feedback on using the game, e-books and AR-based learning materials for learning?). In general, most students enjoyed learning with the game-based learning materials, and stated that they did not feel stressed even when they could not answer the questions since it was a game. They also reflected that the practice in the game could be transformed into hands-on activities to help them complete the scientific toy. However, it was noticed that some students felt that it was challenging to finish the work.

For the feedback from GA, the students gave positive feedback on the media elements of the learning materials, and revealed that the text and figures in the materials helped them to doubly confirm the processes of the learning information. Some students did a quick review of the content and once they got stuck on the processes they would watch the animation video to overcome the problem step by step. Some other students tended to discuss with their peers first, and if they still could not solve the problems, then they would follow the instruction of the materials in detail to complete the work.

The feedback from GB indicated that it was helpful to have AR-based guidance which provided instruction by showing colored blocks on the physical objects to facilitate assembly of the elements to complete the science toy. However, it was also noted that some students reflected that since it took time to scan and get the content from the AR-based materials, they would do the work on their own sometimes. Besides, the students suggested that when the AR triggers were removed from the lens of the learning devices, the learning information disappeared. The design of the mechanism of the AR content could be further improved.

In general, the feedback on using the AR-based and e-book materials indicated that the learners used the learning materials for the hands-on activity in various ways. However, it was noticed that one or two students stated that they would not think before doing the project but just follow the steps in the instructions if the materials provided them with very detailed guidance such as step-by-step instruction. The learners also shared the feedback of using the e-learning materials for project-based learning, stating that they needed teachers' aid before, but with the assistance of the learning materials they could complete the task independently. They were therefore able to complete the project at their own pace.

Table 5. The qualitative feedback from the learners

Items	Students' feedback
Impressive elements in the materials: Text, figures and video	• The figures in the book help me to understand the connection of the Circuit. GA-1-11 • The textual information is very clear. The figures in the content helped me to know the correct position of elements and the animation helped me to know the processes of operation. GA-1-11 • The textual information is better than the figures. GB-2-21
Impressive elements in the materials: 3D objects	• The 3D objects in the content helped me to know how to assemble the scientific toy. GB-1-3 • The 3D objects in the content showed clearly the steps of assembly. GB-2-4 • The figures and animation in the content helped me to know the correct position of the elements, and the 3D objects showed the back of the elements GB-1-8 • The 3D objects showed the information more clearly. GB-1-11 • The textual information helped me to know the steps, and with the 3D objects to present the information, it helped me to understand better. GB-1-20
Feedback on the game-based learning materials	• I understand the concepts more, and I could use the learned skills in the hands-on activity. GA-1-26 • If I did not practice in the game, I might not have finished the hands-on work. I liked the game-based learning, it is more fun. GA-1-24 • I like the game-based learning because it is not stressful and there were hints in the game. GB-1-11 • I think the hands-on activity is harder, because sometimes I did not know how to fix the problems when I got it wrong. GB-1-1
Feedback on the AR-based or e-book learning materials	• The course was not so hard for me, and it was really interesting. GA-1-33 • At first, I was not familiar with the content but I learned a lot in the activity. GB-2-5 • I think that this way (with the help of the e-materials) was good for me. GB-1-19 • With the help of the learning materials, I learned the processes of operation more easily. And it is more effective. GB-1-20 • The activity was very challenging, and I learned a lot. GB-2-23
Suggestions	• The speed of scanning (the AR content) should be faster.GB-1-3 • The learning materials were good, but the speed of scanning should be faster. GB-1-10

5.4. Qualitative feedback from the science teacher

The feedback from the interview of the science teacher was collected and organized to answer the second research questions. The feedback was summarized and organized into three categories. First, the teacher stated that the e-books helped the students to connect the concept knowledge and hands-on processes through textual and image-based information. The students were able to read the text, to think (about the diagram of the circuit), then to complete the hands-on work. The teacher also mentioned that the 3D demonstration of the objects might be needed for assisting students with object assembly (Figure 6). Second, the teacher found that the students' learning motivation and concentration improved with the use of the AR-based materials. The clear step-by-step guidance of the AR information that connected visual learning information with physical objects was very helpful in terms of assisting every student in completing the hands-on work. However, it was also noticed that some learners tended to follow the step-by-step guidance directly without thinking, and the teacher indicated that the mechanical operation might not be a good thing for concept acquisition since they did not know the principles of circuits. Lastly, the teacher suggested that the AR-based content could be further improved through encouraging students to operate the 3D objects of the AR information. The teacher also revealed that since it took more time to prepare the AR environment, such as checking the Wi-Fi connection, better support of the equipment was also needed. Hence, currently, the science teacher had more positive perceptions of using the e-books to support science project-based learning in the classroom.

Figure 6. The learning scenario in the classroom

6. Discussion

After analyzing the quantitative and qualitative data, the discussion is presented according to the research questions (RQs).

6.1. RQ1

To answer the first research question (Are there differences in the learning performance of the students who learn through the game with e-book learning materials and the game with AR-based learning materials?), the results of the quantitative data analysis indicated that generally the GA had higher scores than GB in the hands-on activity according to the descriptive statistics results, but it showed no significant learning difference. Valuable research findings were found from the qualitative results of the interview and the learners' written feedback. The students revealed that the basic science concepts they practiced in the game helped them to complete the hands-on project, and it was noticed that the learning inquiry happened during the hands-on activities. The qualitative data from the interviews indicated that the learners had interaction and cooperation with peers during the processes of producing the scientific electric current toy. In sum, the students had discussions to exchange views and ideas when they met disagreement during the learning activities. These findings are in accordance with the previous research (ChanLin, 2008).

The study also found that the learners exhibited various behaviors when using the support of the e-books and AR-based materials. For example, students with the e-books tended to find out the answers on their own at the beginning of the activity and only when they got stuck would they use the learning materials to find out the answers, but some would like to have peer discussion during the whole hands-on activity and look up the answers when they could not reach agreement. It was also noted that some students with the AR-based learning materials would like to follow the instruction in the materials step-by-step during the activity.

6.2. RQ2

To answer the second research question (What are the students' and teacher's feedback on using the game, e-books and AR-based learning materials for learning?), the questionnaire results showed that both groups were positive about the technology-enhanced learning materials, and the questionnaire results also revealed that the learners with AR supports had better learning confidence in completing the hands-on work. The feedback from the teacher also confirmed the findings that the learners' motivation and learning concentration were enhanced during the whole project-based activity. These findings echo previous research which argued that integrating technology supported students' science learning motivation and engagement (ChanLin, 2008). However, the teacher also stated that well-designed AR-content might restrict students' creativity and hinder their thinking ability since some students followed the instruction and guidance of the AR content directly as soon as they got the materials, without thinking before doing. Despite the e-books not presenting very detailed assembly demonstration, the textual information with the figure as guidance made the learners think first then do the work. This reflection was quite similar to that in Kerawalla's et al. (2006) study. The study argued that the highly interactive technology might sometimes hinder learners' creativity, revealing that AR-based learning content gave students fewer chances to explore questions.

In sum, technology is a good tool to help transform the classroom into a suitable environment for science project-based learning and to achieve the practice of science for knowledge construction. The findings of the current study have revealed that using learning technology for assisting science learning is effective in terms of improving students' active participation as well as science learning motivation. When integrating technology into scientific project-based learning, whether the instruction and detailed guidance provided in the e-learning materials would be helpful or would hinder learners' exploration and creativity ability still needs to be further explored.

7. Conclusion

The study intended to integrate the value of the convenience and accessibility of e-books, the features of interactive demonstration and the combination of visual information with physical AR objects, and the enjoyment brought by game-based learning to create an integrated e-learning model for scaffolding students in acquiring different cognitive knowledge levels through a science project-based activity. Despite the quantitative data not presenting any learning differences between the two groups, the qualitative results showed that the e-learning materials with multimedia content (both the e-books and AR materials) were helpful in scaffolding students while completing the science hands-on activity, and triggered learners' peer discussion to achieve concept agreement.

Following the findings of the current research, I propose a revised e-learning integration model (Figure 7) to further support science project-based learning. In the revised design, the roles of the learning technology tools were adjusted according to various learning objectives of the learning content. The game-based learning environment is designed for factual and conceptual knowledge acquisition, aiming to give students basic concepts of electric circuits, paths, series and parallel, and to recognize circuit diagrams, current, voltage, and resistance. Besides, the revised version of the game-based learning content was improved through adding a new learning unit as scaffolding between basic conceptual learning and practical hands-on activity. The learners can simulate the processes of hands-on activities in the game-based environment and after they are familiar with the steps, they can then transfer their learning to the physical hands-on practice.

Since the core value of project-based activities is not to give students full guidance but to encourage them to think before starting the work and to give them more opportunities to be independent and to engage in self-regulated learning, the e-books with partial AR functionality were used in the revision model to support higher level knowledge acquisition and practical hands-on activities. The revised e-books used text, figures and animation for scientific toy assembly guidance, and the AR model would function only when learners needed further 3D demonstration such as demonstration of objects flipping or location guidance. The aims of the revised e-books tended to combine the advantages of e-books and AR techniques to create more easily accessible e-learning materials to assist students in applying the comprehensive conceptual knowledge at more advanced learning levels.

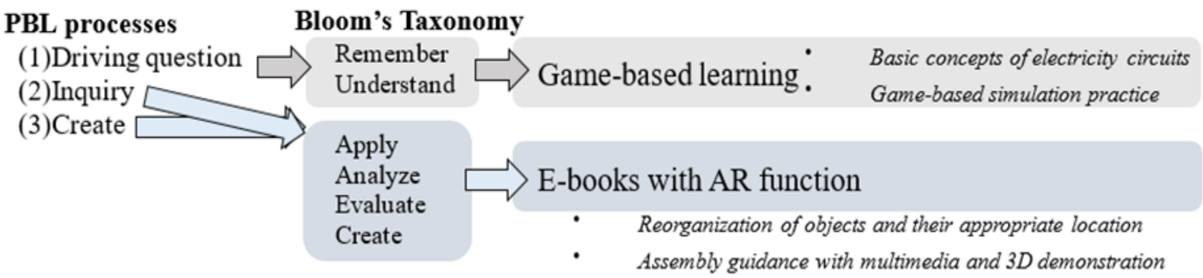


Figure 7. The revised e-learning model to support science project-based learning

The limitations of this study are that firstly the participants were elementary school students and the learning topic is for science education; secondly, the course duration was only 4 weeks due to consideration of the real classroom learning scenario and the teacher's routine instruction. Besides, the current game-based learning materials did not record the students' logging data; hence, the students' learning profiles and records were not analyzed in the current study. The future work will firstly be to revise the design of the e-books according to the findings of this study, and to further improve the game-based learning materials by constructing the database to further record learners' practice processes such as scores for further reference. Furthermore, it is suggested that teachers could consider the roles, affordances and features of various media when integrating them into supporting teaching and learning. Moreover, a longer experiment should be conducted and a delayed test is suggested to further confirm whether the learning technologies truly support science learning in the long run.

Acknowledgement

This research project is jointly funded by the Ministry of Science and Technology in Taiwan, 107-2511-H-032 -004 -MY2 and 109-2511-H-032-003-. The authors would like to thank the Ministry of Science and Technology for their support. The author would also like to thank Jian-Yi Lin, Lien Kuan-Yu and Pei-Jie Ho for the assistance in collaboratively designing and collecting the material and research data.

References

Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for teaching, learning, and assessing: A Revision of Bloom's taxonomy of educational objectives.* New York, NY: Longman.

Barak, M., Ashkar, T., & Dori, Y. J. (2011). Learning science via animated movies: Its effect on students' thinking and motivation. *Computer & Education, 56*(3), 839-846.

Bloom, B., Englehart, M. Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York, NY: Longmans.

ChanLin, L. J. (2008). Technology integration applied to project-based learning in science. *Innovations in Education and Teaching International, 48,* 55-65.

Chauhan, S. (2017). A Meta-analysis of the impact of technology on learning effectiveness of elementary students. *Computers & Education, 105*, 14-30.

Chen, C. H., & Yang, Y. C. (2019). Revisiting the effects of project-based learning on students' academic achievement: A Meta-analysis investigating moderators. *Educational Research Review, 26*, 71-81.

Chen, C. H., Law, V., & Chen, W. Y. (2018) The Effects of peer competition-based science learning game on secondary students' performance, achievement goals, and perceived ability. *Interactive Learning Environments, 26*(2), 235-244.

Chen, C.-H., & Su, C.-Y. (2019). Using the bookroll e-book system to promote self-regulated learning, self-efficacy and academic achievement for university students. *Educational Technology & Society, 22*(4), 33–46.

Chen, C-H., Huang, C.Y., & Chou, Y-Y. (2017). Integrating augmented reality into blended learning for elementary science course. In *Proceedings of the 5th International Conference on Information and Education Technology* (pp. 68-72). New York, NY: ACM.

Chen, L. (2004). Cooperative project-based learning and student's learning styles on web page. *Journal of Educational Technology Systems, 32*(4), 363-375.

Cheng, K.-H., & Tsai, C.-C. (2013). Affordances of augmented reality in science learning: Suggestions for Future Research. *Journal of Science Education and Technology, 22*(4), 449–462.

Collins, A. (1994). Goal-based scenarios and the problem of situated learning: A Commentary on Andersen Consulting's design of goal-based scenarios. *Educational Technology, 34*(9), 30-32.

Dewey, J. (1938). *Experience and education*. New York, NY: MacMillan.

Elliott, S. N., Kratochwill, T. R., Littlefield Cook, J., & Travers, J. (2000). *Educational psychology: Effective teaching, effective learning* (3rd ed.). Boston, MA: McGraw-Hill College.

Embong, A. M., Noor, A. M., Ali, R. M. M., Bakar, Z. A., & Amin, A. R. M. (2012). Teachers' perceptions on the use of e-books as textbooks in the classroom. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, *6*(10), 2638-2644.

Engelhardt, P. V., & Beichner, R. J. (2004). Students' understanding of direct current resistive electrical circuits. *American Journal of Physics, 72*, 98-115.

Ergül, N. R., & Kargın, E. K. (2014). The Effect of project-based learning on students' science success. In *Proceeding of Social and Behavioral Sciences* (Vol. 136, pp. 537-541). Iasi, Romania: Elsevier Ltd.

Eskrootchi, R., & Oskrochi, G. R. (2010). A Study of the efficacy of project-based learning integrated with computer-based simulation - STELLA. *Educational Technology & Society*, *13*(1), 236–245.

Gee, J. P. (2003). What video games have to teach us about learning and literacy. New York, NY: Palgrave/Macmillan.

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

Kerawalla, L., Luckin, R., Seljeflot, S., & Woolard, A. (2006). Making it real: Exploring the potential of augmented reality for teaching primary school science. *Virtual Real, 10*(34), 136-174.

Klopfer, E., Osterweil, S., & Salen, K. (2009). *Moving learning games forward: Obstacles, opportunities, and openness. The education arcade*. Cambridge, MA: Massachusetts Institute of Technology.

Krajcik, J., & Shin, N. (2014). Project-based learning. In R. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (Cambridge Handbooks in Psychology, pp. 275-297). Cambridge, UK: Cambridge University Press.

Krajcik, J. S., & Mun, K. (2014). Promises and challenges of using learning technologies to promote student learning of science. *Handbook of Research on Science Education*, *2*, 337-360.

Lawson, A. E. (1995). *Science teaching and the development of thinking*. Belmont, CA: Watsworth Publishing Company.

Liao, C. Y., Lien, C. J., & Lu, Y. L. (2013). A Study of developing a computer educational game for elementary science teaching on moon and the effectiveness. *Journal of Research in Education Sciences, 21*(3), 317-344

Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., Blunk, M., Crawford, B., Kelley, B., & Meyer, K. M. (1994). Enacting project-based science: Experiences of four middle grade teachers. *Elementary School Journal, 94*, 517-538.

Marx R. W., Blumenfeld P. C., Krajcik J. S., & Solloway, E. (1998). New technologies for teacher professional development. *Teaching and Teacher Education, 14,* 33-52.

Matcha, W., & Rambli, D. R. A. (2013). Exploratory study on collaborative interaction through the use of augmented reality in science learning. *Procedia Computer Science, 25*, 144-153.

Nielsen, B. L., Brandt, H., & Swensen, H. (2016). Augmented reality in science education–affordances for student learning. *Nordic Studies in Science Education, 12*(2), 157-174.

Ozdamli, F., & Turan, B. (2017). Effects of a Technology Supported Project Based Learning (TS-PBL) approach on the success of a mobile application development course and the students' opinions. *TEM Journal-Technology Education Management Informatics, 6*(2), 258–264. doi:10.18421/TEM62- 10

Panasan, M., & Nuangchalerm, P. (2010). Learning outcomes of project-based and inquiry-based learning activities. *Journal of Social Sciences, 6*(2), 252-255.

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist, 50*(4), 258-283.

Remziye, N., & Kargın, E. K. (2014). The Effect of project-based learning on students' science success. *Procedia Social and Behavioral Sciences, 136*, 537-541.

Rosen, Y. (2009). The Effects of an animation-based on-line learning environment on transfer of knowledge and on motivation for science and technology learning. *Journal of Educational Computing Research, 40*(4), 451–467.

Sawyer, R. K. (2004). Creative teaching: Collaborative discussion as disciplined improvisation. *Educational Researcher 33*(2), 12–20.

Shapley, K., Sheehan, D., Maloney, C., & Caranikas-Walker, F. (2011). Effects of technology immersion on middle school students' learning opportunities and achievement. *Journal of Educational Research, 104*, 299-315.

Smeets, D. J. H., & Bus, A. G. (2012). Interactive electronic storybooks for kindergartners to promote vocabulary growth. *Journal of Experimental Child Psychology, 112*(1), 36-55.

Spanovic, S. (2010). Pedagogical aspects of e-textbooks. *OdgojneZnanosti-Educational Sciences, 12*(2), 459–470.

Sumarni, W. (2015). The Strengths and weaknesses of the implementation of project based learning: A Review. *International Journal of Science and Research 4*(3), 478-484.

Sung, T. W., & Wu, T. T. (2018). Learning with e-books and project-based strategy in a community health nursing course. *Computers, Informatics, Nursing, 36*(3), 140-146.

Weng, C., Otanga, S., Weng, A., & Cox, J. (2018). Effects of interactivity in E-textbooks on 7th graders science learning and cognitive load. *Computers & Education, 120,* 172-184.

Yoon, S., Anderson, E., Lin, J., & Elinich, K. (2017). How augmented reality enables conceptual understanding of challenging science content. *Educational Technology & Society, 20*(1), 156-168.

Yoon, S. A., Miller, K., & Richman, T. (2020). Comparative Study of High-Quality Professional Development for High School Biology in a Face-to-Face versus Online Delivery Mode. *Educational Technology & Society*, *23* (3), 68–80.

Comparative Study of High-Quality Professional Development for High School Biology in a Face-to-Face versus Online Delivery Mode

Susan A. Yoon*, Katherine Miller and Thomas Richman
Graduate School of Education University of Pennsylvania, Philadelphia PA, USA // yoonsa@upenn.edu // kmmiller@upenn.edu // trichman@upenn.edu
*Corresponding author

(Submitted April 7, 2020; Revised May 25, 2020; Accepted August 10, 2020)

ABSTRACT: Online professional development (PD) can support broader accessibility than traditional face-to-face PD. However online delivery presents challenges for characteristics of high-quality PD, such as collaborative knowledge building and community development, that have proven positive outcomes in face-to-face modes. A few comparative studies have demonstrated equivalent outcomes when PD activities have been translated from a successful face-to-face implementation to an online format. This study investigates whether an online version of PD for high school biology teachers on using computer-supported complex systems curriculum and instruction can achieve the same high impact as the face-to-face version. We describe changes in design decisions to accommodate the online mode and measure impact on teachers' perceptions of their experiences and student outcomes. The results show positive teacher perceptions in both PD formats and roughly equal student outcomes. However, teachers articulated other benefits to online activities that indicate opportunities for improved access to high-quality PD.

Keywords: Online professional development, Complex systems, Comparative study

1. Introduction

Research has shown that high-quality PD can improve teachers' practice and have a positive impact on student outcomes (e.g., Desimone & Garet, 2015; Fischer et al., 2018). However, for many teachers who are geographically isolated, access to high-quality PD is a primary limiting factor in their professional growth and ability to address the changing landscape of science education (Peltola et al., 2017; Wei et al., 2009; Wilson, 2013). For these teachers, face-to-face PD is often not an option. As Hill (2015) and others have noted, traditional face-to-face PD is expensive and limited in its ability to scale. Offering PD online is a rapidly growing option that can support broader accessibility—anywhere, anytime—at lower costs (Dede et al., 2019; U.S. Department of Education, 2010). Yet, despite its promise, there are challenges in designing high-quality online PD that incorporates known best practices, such social learning and collaborative knowledge building, that have proven effective in face-to-face PD (Dede, 2009; Moon et al., 2014). A report by the U.S. Department of Education (2010) suggested that online PD can provide a more accessible option if it can be shown to be as effective as face-to-face PD. Some studies have demonstrated (e.g., Fishman et al., 2013) that PD offered online can be as successful as face-to-face modes in promoting improved teacher knowledge, classroom practice, and student outcomes. However, very few studies have directly compared effectiveness in both formats (Moon et al., 2014; U.S. Department of Education, 2010). Our study undertakes such a comparison.

From 2010 to 2014 we developed and delivered face-to-face PD to high school biology teachers using computer-supported complex systems curriculum and instruction. The PD was built on known characteristics of high-quality PD for science teachers (reviewed below). Findings from several studies of the PD (e.g., Yoon et al., 2017a; Yoon et al., 2016) show that teachers rated it highly, saw great curricular utility in supplementing and revising their practices, and demonstrated improved student participation and learning outcomes. The success of this project encouraged us to consider how to scale this work to reach more teachers. We consulted recent reports on the lack of high-quality teacher PD that implicated time and space as issues related to scale (Merritt, 2016). This led to a plan to move the PD entirely online. We chose edX as the platform, due to its mechanisms for reaching large numbers. In recreating the PD on edX, we attempted to stay true to the objectives and basic structure of the face-to-face PD while including adaptations in line with best practices for online learning (e.g., Booth 2012; Dede et al., 2019; Yuan & Kim, 2014). In the research reported here, we compare the results from the face-to-face PD to a pilot run of the online course.

2. Theoretical considerations

2.1. Designing high-quality PD

When determining what characteristics constitute high-quality PD for science teachers, some common practices emerge from the literature (e.g., Desimone, 2009; Gerard et al., 2011). These practices include a focus on relevant and useful content, eliciting and capitalizing on teachers' knowledge and skills, engaging teachers in active learning, allowing teachers time and space for reflection, and supporting collaboration. A review of studies focused on PD qualities that support technology-enhanced inquiry instruction in science by Gerard and colleagues (2011) also pointed out the importance of mentoring and access to expertise through facilitators and more senior teachers. The importance of building collaborative teacher communities as a way to provide ongoing PD has been written about for more than 2 decades (e.g., Cochran-Smith & Lytle, 1999; Lieberman, 1995). More recently, we have come to understand that providing opportunities for teachers to build social capital (i.e., the ability to acquire resources from others) is just as essential as supporting the development of teachers' human capital (i.e., skills and knowledge possessed by an individual), especially when resources and experience are limited (Yoon et al., 2017b). For online environments, scaffolded social forums where teachers can interact with one another and build on each other's knowledge and expertise are needed (Zeichner & Liston, 2014). Booth (2012) discussed the value of open exchanges of ideas, experiences, and resources in building trust in online professional learning communities. In setting a research agenda for more online teacher professional development research, Dede and colleagues (2009) outlined a number of additional benefits that include opportunities for reflection and greater participation by teachers who may normally remain silent in face-to-face modes. However, as that study suggested, more research is needed to determine whether and how online PD is a viable alternative to face-to-face PD (Dede, 2009; U.S. Department of Education, 2010).

2.2. What we know about the benefits of online PD

In this section, we provide more details about why online PD for teachers is beneficial. First, it allows for greater geographical diversity among participants and reduces transportation time and costs, as teachers can access online PD courses from anywhere (Bates et al., 2016). Specifically, asynchronous online PD provides learners with the ability to self-pace and access materials at flexible hours. Evidence shows this time flexibility is particularly appealing to PD participants, especially when teaching schedules do not allow for attendance in face-to-face programs (Parsons et al., 2019). Online PD also has the potential to reduce the overall cost of PD implementation when online resources are used repeatedly and participants are geographically dispersed (Fishman et al., 2013). Other researchers have hypothesized that online PD offers a natural avenue for developing and supporting online teacher communities in the longer term, which can help teachers continually reflect on and adapt their teaching practices (e.g., Frumin et al., 2018). Dede and colleagues (2019) and others have discussed the fact that online PD seems well-positioned to take advantage of many of the benefits brought about by learning engineering and other advancements in online educational platforms. Parsons and colleagues (2019) conducted a survey with 258 teachers across 41 U.S. states and found that the majority of respondents had participated in online PD opportunities.

Previous studies and the aforementioned potential notwithstanding, there still remains the issue of limited empirical research directly comparing the outcomes of online versus face-to-face PD in order to show that online PD can be as effective for teachers and ultimately for student learning. Fishman and colleagues (2013) used a randomized experiment to examine the differences in both secondary school teacher and student learning outcomes resulting from a PD delivered face-to-face and asynchronously online. Teachers from both PD groups exhibited similar increases in content knowledge and feelings of self-efficacy, with only minor differences in changes to classroom practices. Student performance also increased comparably across both groups. However, this study included an extensive 16-hour in-person orientation to situate teachers in the program. Given the geographically diverse nature of many online PD programs, we believe these results may not reflect a fully asynchronous PD delivery, especially when the PD content includes the introduction of novel technologies to teachers. Additionally, despite the promising results of Fishman et al. (2013), Moon and colleagues (2014) cautioned against overgeneralizing from this study and recommended additional research into the nuanced trade-offs between online and face-to-face PD. Other studies examining online versus face-to-face adult learning programs in professional and academic settings have suggested that outcomes from online PD could compare favorably with face-to-face PD, though these studies do not include practicing teachers and also lack the fully-asynchronous component (Kissau, 2015; Olivet et al., 2016). Further comparative research in explicitly online and asynchronous PD is needed. Our study asks the following research question: To what extent are asynchronous online PD experiences comparable to face-to-face experiences, in relation to teachers' perceptions and students' learning outcomes?

3. Methods

To address the research question, we used a comparative mixed methods design (Maxwell & Loomis, 2003). It is comparative in that we analyzed differences in teachers' perceptions of the PD delivery and subsequent student learning outcomes when teachers accessed the same curriculum through a face-to-face versus an online asynchronous PD platform. It is a mixed methods design in that we integrated quantitative and qualitative data analyses to reveal differences in patterns and outcomes with information that explains why or how those patterns and outcomes may exist in the data.

3.1. Context

This study is part of a long-standing program of research funded by the U.S. National Science Foundation that undertakes the design and dissemination of a curriculum to teach common topics in high school biology through complex systems modeling (Yoon et al., 2016; Yoon et al., 2017a; Yoon et al., 2018). The curriculum includes five units on the topics of Genetics, Evolution, Ecology, the Human Body, and Animal Systems. They entail working with an agent-based simulation tool and include experiences in core scientific practices as outlined in the *Next Generation Science Standards* (NGSS), such as analyzing and interpreting data, engaging in argument from evidence, and obtaining, evaluating, and communicating knowledge claims. Students normally work in groups of two to complete the units, each of which take 2 to 3 days to complete. The design of the curriculum was conducted alongside the PD activities, described in the following section.

3.1.1. The face-to-face PD

The face-to-face version of the PD ran for two iterations (1.1 and 1.2) between 2012 and 2014. We provide a brief summary here (for further details see Yoon et al., 2016; Yoon et al., 2017a). The participants attended 80 hours of programming over 2 consecutive years, with a 30-hour week-long summer workshop during August of each year and 10 additional hours in weekend workshops during the school year. The PD was designed to align with the following known characteristics of high-quality PD (Desimone, 2009): (a) *Addressing current content:* The biology topics selected for each unit constituted common topics that are mandated in the high school biology curriculum and local, state, and national standards. (b) *Active learning opportunities:* We focused especially on active learning as a way to support teachers in learning and implementing new technologies, a known challenge for teachers. During each summer of the PD, teachers completed the units from the curriculum in pairs to experience the curriculum as a learner and participated in additional activities that engaged them in reflecting on the experience and brainstorming how to support students in the future classroom implementation. Project facilitators were on hand to help teachers make sense of the simulations and activities, particularly when learning how to program. (c) *Coherence with professional demands:* We worked hard to scaffold experiences in several newer conceptions of scientific content outlined in the NGSS, such as integrating computer programming, scientific modeling and argumentation, and systems perspectives. This work included understanding individual teacher developmental trajectories in terms of their knowledge and experience with these different topics and taking into consideration the context within which they taught. We created adapted sessions to meet the teachers where they were in their development and needs to differentiate the curriculum for their student populations (e.g., one teacher taught in a school with a high English Language Learner population). (d) *Extended duration:* The PD took place both during the summer and school year to provide ongoing support and continuity as they implemented the project in their classrooms. Teachers also participated in a second year of PD. (e) *Collective participation:* We supported collaboration by working with specific content and age-level teachers in high school biology, several of whom were from the same schools, which allowed them opportunities to address coherence with professional demands. During each school year, participants conducted at least four of the curriculum units with their students.

The first year—2012–2013, which we refer to as Iteration 1.1—focused on the content and pedagogy of the curriculum, with participating teachers acting as co-designers in iteratively improving the simulations and supporting student and teacher materials. Based on feedback from teacher surveys and interviews, the second year—2013–2014, which we refer to as Iteration 1.2—included greater focus on increasing collaboration and relationship building among the teachers. (For more details about both years of PD activities, see Yoon et al., 2017a; Yoon, 2018.) Iterations 1.1 and 1.2 were used as comparison to the online implementation (Iteration 2.0) because they provided information about how to improve the curriculum materials and how to work with teachers as a community, which are both elements of high-quality PD.

3.1.2. The online PD

The online version of the PD (Iteration 2.0) ran July 2018 through June 2019. It was designed to create a space that, similar to the face-to-face PD, fostered social relationships as a way to engage in active sense making. Based on results and feedback from Iteration 1.0 (Yoon et al., 2017a), we found five categories of supports that the teachers indicated were most important for their learning from the PD. This included the following: (a) *supporting teacher beliefs*—teachers said that over the course of the PD, their confidence and beliefs about the utility of the curricular emphases (e.g., in modeling, argumentation, and systems) grew and was an important factor in encouraging shifts in their teaching practice; (b) *time and experience*—teachers believed that the 2-year PD afforded them more time to hone their expertise with the curriculum and more experience in the classroom in implementation; (c) *hands-on practice and training*—teachers found that the second year was particularly valuable in that they were able to garner more hands-on practice and training in learning how to use and teach with the computational models; (d) *just-in-time supports*—teachers believed that having the facilitation from project team members available to them while they were planning and implementing activities in the classroom was important; and (e) *interaction and building knowledge within the community*—teachers felt that the greater emphasis on sharing ideas and building relationships with other teachers in the PD provided them with more resources and partners to think through problems of practice.

The online PD was designed to provide these supports while offering a modified version of the in-person activities and discussions (including the known characteristics of high quality PD) for fully online delivery. Similar to the face-to-face iterations, the online PD was constructed to take 40 hours to complete and included the other known characteristics of high quality PD. To support teacher beliefs, we built dedicated modules on modeling, argumentation, and systems that provided readings, weblinks, and videos about the importance of each concept in scientific exploration. To provide the needed time and experience to become comfortable with project activities, teachers were able to move through the online PD modules at their own pace. We also believed that the archival nature of the online resources would allow teachers to access them anywhere, anytime, and as often as they needed. To provide hands-on practice and training, detailed walk-through videos with voice-over instruction were recorded to provide teachers with online support to practice the simulations on their own. Additionally, while teachers in the face-to-face PD presented problems of practice from their classroom implementation and had in-person discussions about strategies as a way to build knowledge and support teacher beliefs, participants in the online PD watched videos of teachers from Iteration 1.0 teaching the units in their classrooms and discussed strategies on a discussion forum that included the experienced teachers as active members. These videos and the presence of and access to experienced project teachers were also meant to function as just-in-time supports for teachers participating in the online PD. Finally, to support teacher interaction and community knowledge building, we created a curated online discussion forum with prompts that were used to seed discussion in the areas of teacher understanding of course content, classroom implementation, and collaboration or the sharing of resources and expertise.

The above modifications were made intentionally to maintain the objectives of the PD, the characteristics of high-quality PD, and the supports that were important to teachers in 1.0. To guide this transition and oversee consistency, three teachers from Iteration 1.0 acted as design collaborators for 2.0. They worked with the research and design team to provide insight on how well the online version compared to the face-to-face version of the PD. The online course consisted of seven modules and was designed to encompass 40 hours of instruction to be completed over a 6-week period. Additionally, three 2-hour synchronous online meet-ups were held during the school year. These meet-ups were designed to support participants in their implementation by providing space for sharing challenges and successes.

3.2. Population

The participants in each iteration of the study were recruited through the researchers' networks and were primarily a sample of convenience, though they represent many diverse classroom contexts.

3.2.1. Face-to-face teacher participants

Iteration 1.0 of the PD included 10 teachers, seven women and three men, from a metropolitan area in the northeastern United States. The years of experience for the teachers ranged from 3.5 to 19 years, with an average of 8 years. The schools they taught at were all public schools but ranged in demographics, with the student body at one school almost entirely White (3% minority) and another 71% ethnic minority. The socioeconomic demographics were also wide ranging, with school measurements of low-income students ranging from 14% to 59%. Across the 10 participating teachers' classrooms, more than 300 students were reached each year, with 354 participating in research in Iteration 1.1 and 361 participating in Iteration 1.2. See previous results for more detailed demographic information on teachers (Yoon et al., 2017a; Yoon, 2018) and schools and students (Yoon et al., 2016; Yoon et al., 2017a) for 1.0.

3.2.2. Online teacher participants

For the online PD, six teachers from five different schools in three metropolitan areas in the northeastern United States participated in the course. The teachers had a range of 3 to 19 years of teaching experience with an average of 8.5 years. The teachers were employed mainly in public schools with the exception of one independent school. Two of the teachers taught at the same school, and the other four were each at one of the other schools. The demographics of these schools also demonstrated diversity. The most diverse school had 82% ethnic minority, while the least diverse had 19%. The percentage of low-income students ranged from 62% to less than 5%. Five of the participating teachers agreed to collect data on their students and, with some attrition, resulted in data from 88 participating students for the 2.0 dataset. See Table 1 for further details.

Table 1. Demographics of teachers and schools

Teacher	Years of experience	Number of students	School level characteristics		sample
			% Low income	% Minority	% Minority
1-1	3.5	45	59.0%	71.7%	79.7%
1-2	14	12	59.0%	71.7%	79.7%
1-3	8	22	27.7%	38.9%	43.4%
1-4	12	16	36.0%	NA	31.6%
1-5	5	46	59.0%	71.7%	79.7%
1-6	8	23	NA	NA	NA
1-7	5	48	27.7%	38.9%	43.4%
1-8	6	63	14.8%	29.9%	42.2%
1-9	19	21	14.2%	3.2%	20.0%
1-10	5	65	18.0%	36.7%	38.2%
Average	8.6	36.1	35.6%	42.7%	40.4%
Total		361			
2-1	3	29	NA	19%	10.3%
2-2	11	16	4.7%	24%	50.0%
2-3	19	14	30.4%	41%	42.9%
2-4	7	14	63.0%	82%	92.8%
2-5	4	NA	NA	NA	NA
2-6	7	15	4.7%	24%	57.1%
Average	8.5	24.8	25.7%	38%	50.6%
Total		88			

3.2.3. Student participants

Student populations were similar across the three iterations, with more female than male participants, and with a skew towards lower grades. See Table 2 for further details.

Table 2. Demographics of students

Student characteristics	Iteration		
	1.1 (2012-2013)	1.2 (2013-2014	2.1 (2018-2019)
Number of students (n)	363	361	88
Gender			
Male	160 (44.1%)	157 (43.4%)	34 (38.6%)
Female	177 (48.7%)	198 (54.9%)	50 (56.8%)
Grade			
9^{th}	137 (37.7%)	170 (47.1%)	36 (40.9%)
10^{th}	71 (19.6%)	75 (20.8%)	30 (34.1%)
11^{th}	61 (16.8%)	92 (25.5%)	15 (17.0%)
12^{th}	5 (1.4%)	14 (3.9%)	6 (6.8%)

3.3. Data sources

To investigate our research question, data from four sources were collected from each of the three iterations of the PD: teacher post-course satisfaction surveys, teacher post-course and post-implementation interviews, student pre- and post-implementation content knowledge surveys, and student pre- and post-implementation classroom experience surveys. In this paper, we present data on the teacher post-course satisfaction surveys and student content knowledge and experience surveys from both the face-to-face and online PD versions. We present teacher-interview data only from the most recent online PD version, which was used to elaborate on outcomes from teacher- and student-survey data.

Upon completing the PD summer course, teachers were administered a PD satisfaction survey to measure their perceptions of the PD. This comprised 18 five-point Likert-scale questions (1 = strongly disagree to 5 = strongly agree) that probed their experiences with the course resources in the areas of overall course satisfaction (e.g., “The course covered topics that are relevant to the grade(s) I teach”); module construction and delivery (e.g., “The modules actively engaged those in attendance”); and usability of materials in specific teaching activities (e.g., “The student worksheets given out during the course will be useful in my teaching”). This survey has been used in previously published studies (e.g., Yoon et al., 2015) to ascertain teachers’ opinions on the extent to which the PD was enjoyable and had utility in their instruction. Studies have also shown that teachers perceptions of the effectiveness of participation in online teacher professional learning communities is strongly mediated by their sense of learning satisfaction (e.g., Tsai, 2012).

Individual interviews for the online PD participants were conducted with teachers at the end of the summer course and again at the end of the school year. Post-course surveys used semi-structured questions to gather information about teachers’ experience in the course in terms of the online format (e.g., “Were there specific topics you felt were better delivered in an online format and others face-to-face?”), social interactions with peers (e.g., “To what extent did you find the discussion aspect important to your learning?”), and support from the project team (e.g., “Do you feel you received adequate support from the team and the prior project teachers?”). Post-implementation interviews were conducted with teachers to gather information about how they implemented the curriculum in their classrooms and how participation in the course prepared them to do so (e.g., “What aspect of the online course did you feel best prepared you for teaching?”). Additionally, they were asked about the usefulness of the online meet-ups (e.g., “What was the most helpful aspect of the online video meetups?”) and the support from the project team that continued into the school year (e.g., “Do you feel you received adequate support from the project team during the school year?”). To understand teachers’ perceptions about benefits to student learning, we also asked the following question: “Discuss some of the benefits for your students you saw with particular modules.” Individual interview lengths ranged from 22 to 42 minutes, and the audio-recorded interviews were transcribed.

Students in both the face-to-face and online versions of the PD course completed two surveys pre-implementation and two surveys post-implementation. Though the surveys contained the same questions, they were administered 9 months apart, to mitigate the effects of item exposure. The first survey testing content knowledge consisted of 14 multiple choice questions about biology content (e.g., “There are many different enzymes located in the cytoplasm of a single cell. How is a specific enzyme able to catalyze a specific reaction?”) and one open-ended question about complex systems (e.g., “Imagine a flock of geese arriving in a park in Philadelphia, where geese have not lived before. Describe how the addition of these geese to the park may affect the ecosystem over time. Consider both the living and nonliving parts of the ecosystem.”). A classroom experience survey was also administered that consisted of 44 five-point Likert-scale questions (1 =

strongly disagree to 5 = strongly agree) that probed students' experiences with instructional strategies in the classroom that were targeted for improvement in the PD (e.g., "I often work together with other students to learn about science; I use computer technologies to share information about science").

3.4. Data analysis

We identified overarching trends and triangulated findings between the four data sources. For the satisfaction surveys, average Likert-scale responses were calculated for all 18 items and then aggregated in the three areas of overall course satisfaction, module construction and delivery, and usability of materials in specific teaching activities. The averages for all three iterations of the PD were compared for equivalence in the three areas across the 3 years. The teacher interview transcriptions were qualitatively mined for comments and insights that offered support for the findings from the surveys as well as insights into how well we were able to meet the needs of teachers in the five categories of PD support that Iteration 1.0 teachers believed were the most important to address (i.e., supporting teacher beliefs, time and experience, hands-on practice and training, just-in-time supports, and interaction and building knowledge within the community).

Scores of the student multiple choice content questions were compared from pre- to post-implementation for each year using a paired *t*-test to see if the average scores changed from the beginning to the end of the curriculum implementation. To understand the amount of impact, we calculated the effect size (Cohen's *d*). Responses to the open-ended complex systems question were scored on a scale of 1 (not complex) to 3 (completely complex) for each of four different dimensions of complex systems understanding, which are listed in Table 3. The pre- and post-responses were combined and two researchers on the project each coded 40 (23%) responses, achieving a Cohen's Kappa interrater reliability score of 0.87. The remaining responses were coded by one of the researchers. These components were derived from earlier research (Yoon, 2008; Yoon, 2011; Jacobson et al., 2011; Pavard & Dugdale, 2000). Aggregate scores on this assessment ranged from 4 to 12.

Table 3. Categories of complex systems components

Complex systems components	Descriptions
Predictability	The emphasis is on the predictability of the effects caused by the agent in question. In a complex framework, it is impossible to anticipate precisely the behavior of the system. This is because the actions of agents cannot be predicted (as random forces or chance factors can affect an agent's actions) even if we know the rules or characteristics of the agent.
Processes	Processes refer to the dynamism of the mechanisms that underlie the phenomena (i.e., how the system works or is thought to work). In a complex systems framework, there is no definite beginning and end to the activity. System processes are ongoing and dynamic.
Order	The focus is the organization of the system or phenomenon as centralized or decentralized. In a complex systems framework, control is decentralized and distributed to multiple parts or agents. Order in the system is self-organized or "bottom-up" and emerges spontaneously.
Emergence and scale	Emergence refers to the phenomenon where the complex entity manifests properties that exceed the summed traits and capacities of individual components. In other words, these complex patterns simply emerge from the simpler, interdependent interactions among the components. In a complex system, because parts or agents are interdependent in multiple ways, an action (small or large) that is imposed on the system may have large and far-reaching consequences on the numerous parts and agents of the system. This may in turn result in large-scale change and evolution.

For more information about the data and reliability analyses on the complex systems coding for Iteration 1.0, see Yoon et al., 2017a. The 2.0 data was coded by two researchers who were trained on the coding manual by a researcher who conducted the analysis of the 1.0 data. The two researchers then coded 23% of the responses (30 out of 132 pre- and post-responses) independently. An interrater reliability of .88 was achieved. One of the researchers then coded the remainder of the responses. The results of this coding were then compared using a paired *t*-test of total scores.

For the classroom experience survey, averages of Likert-scale responses on the pre- and post-surveys and analyzed using paired *t*-tests to determine whether average scores changed.

4. Findings

Findings from the online course for both teacher perceptions and student outcomes showed similar positive increases in usability, interest, and growth to the face-to-face versions of the PD. We provide more details of these findings below.

4.1. Teacher perceptions of online delivery indicate opportunity for accessible PD

Both the satisfaction surveys and the interviews highlighted positive perceptions that teachers had of the online course. The satisfaction surveys showed comparable results to the face-to-face course and the interviews provided reinforcement for the focus on the five areas of support.

4.1.1. Positive satisfaction survey results

Findings from the satisfaction survey showed that teachers in the online PD had a very positive experience, rating all 18 Likert-scale items on average between 4.5 and 5. Aggregate averages in the areas of overall course satisfaction, module construction and delivery, and usability of materials were 4.60, 4.69, and 4.70, respectively. These findings were roughly comparable to those found in Iteration 1.1 of the face-to-face PD (4.73, 4.78, 4.66) but slightly less than Iteration 1.2 (4.95, 4.98, 4.84). In the category of usability of materials, the online teachers (Iteration 2.0) rated items as higher than the teachers in the first year of face-to-face PD (4.70 for online, compared to 4.66 for Iteration 1.1 of face-to-face), which is significant because those items asked whether the specific teacher resources (learned about and acquired online) would be used and/or helpful to their instruction. Teacher interviews also signaled the same positive feelings about the usability of the online PD resources. For example, multiple teachers in the online PD said that the curriculum fit well into their existing Biology course. One teacher, Amber, said:

> I already do these topics. So, on a day where I might have done a different activity, I can just do this instead. So, I love that part of it. I feel like this is actually doable.

While commenting on the computer programming aspect of the curriculum, another teacher, Fran, mirrored Amber's enthusiasm, saying that the new curriculum might facilitate learning better than her old way of teaching:

> This is like just stuff that fits exactly into my curriculum . . . I'm getting pretty excited about kind of changing my curriculum in a way that hopefully facilitates the learning a little better but also kind of filling that need for computer science.

4.1.2. Success of teacher supports

In all five of the categories of PD supports that we intentionally attended to in the design of the online course, we also saw overwhelmingly positive responses. In the category of *supporting teacher beliefs*, there was near unanimity in their interests and beliefs about the utility of modeling, argumentation, and systems in the science classroom based on their experience with the course. For example, Fran said the following about the importance of systems understanding:

> It was a valuable experience for the kids to get their hands on systems modeling. Because that's like I see a lot of . . . I just feel that that makes people wiser to understand systems. And I also think it's important for job prospects later. I think even if kids don't go into science, knowing that like the world functions as like interconnected systems is just like a valuable way of seeing things. So, I thought that was huge.

Teachers also felt that their confidence and skills in delivering these curricular concepts improved, and this was particularly important for teachers whose schools had a similar instructional focus. For example, Amber said:

> Yes, especially argumentation. We as a department have actually been focusing on that a lot in the past month or so, and it's actually a goal in the department to flip our curriculum into more claim, evidence, reasoning, and inquiry. So . . . we did our ecology unit this year using inquiry and argumentation, and [this project] helped me to be more comfortable with that, because it's not easy when you're used to teaching something a

certain way, and then you do it a different way. You let the kids come up with it, and let the kids figure it out through their evidence and reasoning. So, because I did [this project], I was much more comfortable with that switch and doing it. I feel like it really helped prep me, and it will continue to help prep me because we're going to do this with all of our units.

The video footage of 1.0 teachers implementing project activities proved to be the most valuable resource in addressing the category of *time and experience* for the online teachers who felt that being able to visualize events in the classroom was an enormous help in anticipating student learning challenges. Being able to witness real classrooms and teachers using the activities also provided insights into teaching experiences that boosted their confidence. These thoughts are encapsulated in the following comment from Melanie:

> Not only was I learning something but I also liked the videos, and even though we only had to watch snippets of the teachers implementing the modules in the classroom, just seeing someone else all doing it, was like so helpful to me to figure out how I was going to implement it. Because, if I just had the modules and I just had the materials like, there's a big leap between just having a worksheet and knowing how to guide the students through and having the videos where the teachers were reflecting on what went well or what went better, I think. I felt way more confident than just like getting the material that I could actually implement it in my classroom, so I thought that was really helpful. I like the way it's built into like, giving background information and then the challenge like you trying to like model yourself and getting frustrated and trying again like putting yourself in the mind of the student before jumping into, okay, this is how teachers implement it in their classrooms. So, it kind of brought you through the student experience and then the teacher experience. I thought that was great.

For the category of *hands-on practice and training*, teachers appreciated the self-paced nature that the online PD afforded. Amber commented:

> Cause they said go in this as a student, which I felt was really important to see what they would see, and I honestly don't know if I would have time to do that during the school year.

Similarly, Melanie discussed the utility of having time to reflect on her practice after she was able to work at her own pace:

> I would go through the module myself, do the whole worksheet myself that took, you know, maybe like an hour and then an hour or two just thinking about like what did I want to directly introduce. Did I want to make like a slide to show them? Did I want to like, you know, put a little bit of the simulation up on the board just to kind of direct them?

In the category of *just-in-time supports*, all teachers commented positively on being able to access the course resources just before they taught the unit in their classroom, as evidenced in these comments by Elizabeth and Amber:

> Well, before I taught the module, each module I went back and looked at the online aspect of the course that is about that module and seeing the different teachers teach it, which is really helpful.

> I went back . . . because I wanted to refresh myself on a couple of things . . . So, I went back again and re-looked at some of that stuff to kind of remind myself of what I did when I went through it. To be honest with you, I probably should've spent a little more time. I did a quick review, like, "Oh, yeah. I remember doing this. Yep. Yep."

Regarding the category of *interaction and building knowledge within the community*, teachers were again positive about our efforts to scaffold opportunities for sharing. Here is what Fran said about the school-year synchronous online meet-ups:

> It was helpful to hear about the different challenges that people had with implementation so that I could like think about how I might solve those. It just, it's always a good learning opportunity for me to hear about other teacher's experiences, because it helps me reflect on my own practice and when I get to like share ideas, even if I don't share an idea out loud, if I'm able to generate an idea of like, how might I have solved that problem? It is a good practice for me as a teacher to be improving my skills.

Elizabeth similarly commented about the necessity to join the meet-ups to hear about implementation issues that she would not otherwise have known about given that she was the only person implementing the project in her school:

> He could offer adjustments based on solutions you've found or just make a mental note like, "Okay, this is challenging. So, I have to think about this beforehand even if I'm not doing the unit for like two months." I'm just using each other as resources because nobody else at my school obviously is doing it. So, it was really helpful to be like, "Oh yeah, this works really well. I'm going to ask this particular question," or "When the groups were grouped in this particular way that was helpful." So, all those kinds of tools of practice are really good.

4.2. Similar student outcomes for face-to-face and online PD delivery modes

Student outcomes were measured in three areas through two survey tools. The results from Iteration 2.0 are compared here to the results from the first two iterations, which can also be found in more detail in Yoon et al. (2017a). Students' content knowledge in biology improved significantly for all three iterations of the PD (Table 4) with medium effect sizes for the two iterations of the face-to-face PD (d = 0.56 and 0.67) and a small to medium effect size (d = 0.36) for the online PD. Students' understanding of complex systems also improved significantly for all three iterations of the PD (Table 5). Here, the online PD showed a greater effect size (d = 0.38) than the first year of the face-to-face PD, which had only a small effect (d = 0.19).

With respect to the student classroom experience survey measuring students' perceptions of their learning, there was a significant positive increase in all three iterations of the PD (Table 6) with the online PD showing a greater effect size than Iteration 1.1 of the face-to-face PD but not as high as Iteration 1.2. Of note, the data shows similar pre-test averages and standard deviations for all three measures across the Iterations 1.1, 1.2, and 2.0, which suggests some level of equivalence for the populations being measured. These results show that online delivery of PD can have comparable student outcomes to PD delivered face-to-face.

Table 4. Student biology content knowledge results across three project iterations

Iteration	Pre-test Avg* (*SD*)	Post-test Avg* (*SD*)	Difference	*t* value	*df*	*p* value	Cohen's *d*
1.1	6.71 (2.31)	8.13 (2.78)	1.42	10.73	362	< .001	0.56
1.2	7.67 (2.36)	9.43 (2.47)	1.76	12.50	345	< .001	0.67
2.0	7.63 (2.46)	8.50 (2.42)	0.87	3.69	87	< .001	0.36

Note. *Scores are out of 14.

Table 5. Student complex system content knowledge results across three iterations

Iteration	Pre-test Avg* (*SD*)	Post-test Avg* (*SD*)	Difference	*t* value	*df*	*p* value	Cohen's *d*
1.1	6.18 (1.48)	6.51 (1.49)	0.33	3.51	353	< .001	0.19
1.2	5.80 (1.23)	6.79 (1.29)	0.99	12.26	360	< .001	0.65
2.0	5.95 (1.25)	6.51 (1.51)	0.56	2.62	63	< .05	0.38

Note. *Scores are out of 12.

Table 6. Student classroom experience survey results across three iterations

Iteration	Pre-test Avg* (*SD*)	Post-test Avg* (*SD*)	Difference	*t* value	*df*	*p* value	Cohen's *d*
1.1	3.44 (0.48)	3.63 (0.56)	0.19	6.86	320	< .001	0.36
1.2	3.51 (0.44)	3.80 (0.45)	0.29	11.54	309	< .001	0.67
2.0	3.77 (0.41)	3.94 (0.42)	0.17	3.35	80	< .01	0.41

Note. *Scores are out of 5.

The results from the surveys that show significant increases in both student content knowledge and perceptions of science are supported by data from the teacher post-implementation surveys. A common theme across comments from the online PD teachers (Iteration 2.0) was the way that students understood key concepts about of complex systems. Amber commented on how seeing random variation occurring in the simulations provided a different perspective for students, which allowed for deeper understanding of the content:

> They seemed to really grasp the fact of how random some of these events are, even though they do not seem random, in terms of the genes turning on and off. And tying in the complex systems, that's a topic that we go quickly over, and they just say, "Oh, they go on and off," but they do not really know it happens or why it happens. So, this [simulation] actually showed them how and why that happens.

Another theme among teacher comments about student learning was the value of scientific argumentation. Sandy discussed how her experience with the online module on argumentation permeated all aspects of her teaching, which in turn shifted student behavior toward the scientific practice of identifying and using evidence, as the following comment illustrates:

> Every single time, even for a small thing that's going on, we just point out, okay use your evidence and by now students really welcome it. They use it automatically. This is what's happening, this is my reasoning, and this is my evidence.

5. Discussion

The goal of this project was to examine the viability of scaling access to high-quality PD at low to no cost by adapting a successful face-to-face PD initiative to an online delivery mode. Some of the strongest barriers to teachers' professional growth are access to expertise and resources and time constraints (Merrit, 2016; Peltola et al., 2017). By moving PD online, scaling costs can be cut while access for those with time and geographic constraints can be increased. Though some studies (e.g., Fishman et al., 2013; Kissau, 2015) have suggested that online PD can be just as effective as face-to-face PD, more direct comparison research is needed to support these findings (Moon et al., 2014). This study adds to this research with additional evidence that online PD can be as effective as face-to-face PD. Previous research has shown that more time and training with a curriculum lead to stronger and more sustained change and that the ideal duration of PD is 2 years (Gerard et al., 2011; Yoon et al., 2017a). This can potentially explain the higher outcomes from Iteration 1.2 of the face-to-face PD. The outcomes from the online delivery were comparable to, and in some cases even better than, the outcomes from Iteration 1.1 of the face-to-face PD. This suggests that though extended time is beneficial to teachers, similar results can be achieved in the same amount of time whether the PD is delivered online or face-to-face.

The question then becomes how online PD can be designed as a viable alternative to face-to-face PD (Dede, 2009; U.S. Department of Education, 2010). In this study, we have detailed a number of design decisions that we believe supported the comparable satisfaction results from teachers and the learning and experience results from students in both PD modalities. First, our data showed that attending to teacher beliefs that were aligned with what Desimone (2009) described as coherence with professional demands was important to the usability of PD activities and resources. For example, two teachers in our study (Amber and Sandy) reported that in their schools developing student skills in scientific argumentation was a major focus (as we found in several of our 1.0 and 2.0 schools). This state of affairs greatly encouraged teachers' shifts in pedagogical approaches. Next, our data was replete with comments about the overall success of the videos of 1.0 teachers' implementations. We believe that being able to visualize real classrooms and hear from real teachers who also provided metalevel information about pedagogical choices sped up the time that it would have normally taken teachers to feel comfortable teaching with the project resources. This, coupled with the ability for teachers to work at their own pace trying out the activities that they were going to ask their students to complete, enabled time for authentic reflection. We discussed research about these design features in the introduction and theoretical considerations (e.g., access to expert teachers; Gerard et al., 2011); affording time for teachers to reflect on their teaching practices (Frumin et al., 2018); and we offer this insight not so much as something new to consider for the delivery of online PD but as part of a combination of design choices that we believe led to our successful project outcomes. Lastly, we hypothesized that the online nature of PD, with the anywhere, anytime accessibility, would be a component in the ability to deliver just-in-time supports and opportunities for teacher interaction and building knowledge within the community. Our data suggest that this was a good hypothesis given that all teachers mentioned that they returned to the online course to refresh their understanding before teaching with the project activities and all teachers who attended the synchronous online meet-ups found high utility in participation. It is important to note that these teachers were geographically dispersed in different parts of the country, which was a central consideration in undertaking this research with the eventual goal of providing access to high-quality PD for teachers who don't normally have access (Peltola et al., 2017). In other forthcoming research, we discuss the success of additional design efforts that we made in encouraging collaborative discourse in the discussion forum.

A limitation of this study is the relatively small sample size of students from classes of teachers who participated in the online PD. It could be that the sample was biased in some way. However, one of the primary benefits of an online delivery mode is its ability to reach a larger number of teachers. In the summer of 2019, the course was launched on edX, enrolling 260 teachers in 20 countries and 17 U.S. states, with a 16% completion rate (41 teachers). Additional data is being collected on a subset of these teachers and their students in order to provide a more robust data set for comparison with the face-to-face data.

Acknowledgment

This research was funded by two US National Science Foundation Discovery Research K-12 grants (DRL #1721003 and DRL #1019228). We would like to acknowledge the following colleagues who have worked on various portions of these projects: Eric Klopfer, Daniel Wendel, Ilana Schoenfeld, Emma Anderson, Josh Sheldom, Hal Scheintaub, Jessica Koehler, Chad Evans, Miyoung Park, Joyce Lin, Sao-Ee Goh, Murat Oztok, and David Reider.

References

Bates, M. S., Phalen, L., & Moran, C. (2016). Online professional development: A Primer. *Phi Delta Kappan*, *97*(5), 70–73.

Booth, S. E. (2012). Cultivating knowledge sharing and trust in online communities for educators. *Journal of Educational Computing Research, 47*(1), 1–31.

Cochran-Smith, M., & Lytle, S. L. (1999). Chapter 8: Relationships of knowledge and practice: Teacher learning in communities. *Review of Research in Education*, *24*(1), 249–305.

Dede, C., Ketelhut, D. J., Whitehouse, P., Breit, L., & McCloskey, E. M. (2009). A Research agenda for online teacher professional development. *Journal of Teacher Education*, *60*(1), 8–19.

Dede, C., Richards, J., & Saxberg, B. (Eds.). (2019). *Learning engineering for online education: Theoretical contexts and design-based examples*. New York, NY: Routledge, Taylor & Francis Group.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*, 181–199.

Desimone, L., & Garet, M. S. (2015). Best practices in teachers' professional development in the United States. *Psychology, Society and Education, 7*(3), 252–263.

Fischer, C., Fishman, B., Dede, C., Eisenkraft, A., Frumin, K., Foster, B., Lawrenz, F., Levy, A., & McCoy, A. (2018). Investigating relationships between school context, teacher professional development, teaching practices, and student achievement in response to a nationwide science reform. *Teaching and Teacher Education*, *72*, 107–121.

Fishman, B., Konstantopoulos, S., Kubitskey, B. W., Vath, R., Park, G., Johnson, H., & Edelson, D. (2013). Comparing the impact of online and face-to-face professional development in the context of curriculum implementation. *Journal of Teacher Education*, *64*(5), 426–438.

Frumin, K., Dede, C., Foster, B., Lawrenz, F., Eisenkraft, A., Fishman, B., Levy, A., & McCoy, A. (2018). Adapting to large-scale changes in Advanced Placement biology, chemistry, and physics: The Impact of online teacher communities. *International Journal of Science Education*, *40*(4), 397–420.

Gerard, L., Varma, K., Corliss, B., & Linn, M. (2011). Professional development for technology-enhanced inquiry science. *Review of Educational Research*, *81*(3), 408–448.

Hill, H. (2015). Review of *The Mirage: Confronting the hard truth about our quest for teacher development*. Boulder CO: National Education Policy Center.

Jacobson, M. (2001). Problem solving, cognition, and complex systems: Differences between experts and novices. *Complexity*, *6*(3), 41–49.

Kissau, S. (2015). Type of instructional delivery and second language teacher candidate performance: Online versus face-to-face. *Computer Assisted Language Learning*, *28*(6), 513–531.

Lieberman, A. (1995). Practices that support teacher development: Transforming conceptions of professional learning. *The Phi Delta Kappan*, *76*(8), 591–596.

Maxwell, J. A., & Loomis, D. (2003). Mixed method design: an alternative approach. In A. Tashakori & C. Teddlie (Eds), *Handbook of mixed methods in social and behavioral research* (pp. 241–271). Thousand Oaks, CA: Sage.

Merritt, E. G. (2016). Time for teacher learning, planning critical for school reform: Students aren't the only ones who need more time to learn; teachers also need more and better time for learning and for planning. *Phi Delta Kappan*, *98*(4), 31–36.

Moon, J., Passmore, C., Reiser, B. J., & Michaels, S. (2014). Beyond comparisons of online versus face-to-face PD: Commentary in response to Fishman et al., "Comparing the impact of online and face-to-face professional development in the context of curriculum implementation." *Journal of Teacher Education*, *65*(2), 172–176.

Olivet, J., Zerger, S., Greene, R. N., Kenney, R. R., & Herman, D. B. (2016). Online versus face-to-face training of critical time intervention: A Matching cluster randomized trial. *American Journal of Distance Education*, *30*(4), 237–249.

Parsons, S. A., Hutchison, A. C., Hall, L. A., Parsons, A. W., Ives, S. T., & Leggett, A. B. (2019). U.S. teachers' perceptions of online professional development. *Teaching and Teacher Education*, *82*, 33–42.

Pavard, B., & Dugdale, J. (2000). The Contribution of complexity theory to the study of socio-technical cooperative systems. In A. A. Minai, D. Braha, & Y. Bar-Yam (Eds.), *Unifying themes in complex systems* (pp. 39–48). Berlin, Germany: Springer.

Peltola, P., Haynes, E., Clymer, L., McMillan, A., & Williams, H. (2017). *Opportunities for teacher professional development in Oklahoma rural and nonrural schools* (REL 2017–273). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from http://ies.ed.gov/ncee/edlabs.

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

U.S. Department of Education, Office of Planning, Evaluation, and Policy Development. (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies.* Washington, DC: Author.

Wei, R. C., Darling-Hammond, L., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad* (Technical report). Oxford, OH: National Staff Development Council.

Wilson, S. M. (2013). Professional development for science teachers. *Science, 340,* 310–313.

Yoon, S., Anderson, E., Koehler-Yom, J., Sheldon, J., Schoenfeld, I., Wendel, D., Scheintaub, H., Klopfer, E., Oztok, M., & Evans, C. (2015). Design features for computer-supported complex systems learning and teaching in high school science classrooms. *Journal of Research in STEM Education*, *1*(1), 17–30.

Yoon, S. (2018). Mechanisms that couple intentional network rewiring and teacher learning to develop teachers' social capital for implementing computer-supported complex systems curricula. In S. Yoon & K. Baker-Doyle (Eds.), *Networked by design: Interventions for teachers to develop social capital* (pp. 7-23). New York, NY: Routledge Press.

Yoon, S., Goh, S., & Park, M. (2018). Teaching and learning about complex systems in K–12 science education: A Review of empirical studies 1995–2015. *Review of Educational Research*, *88*(2), 285–325.

Yoon, S., Anderson, E., Koehler-Yom, Evans, C., Park, M., J., Sheldon, J., Schoenfeld, I., Wendel, D., Scheintaub, H., & Klopfer, E. (2017a). Teaching about complex systems is no simple matter: Building effective professional development for computer-supported complex systems instruction. *Instructional Science*, *45*(1), 99–121.

Yoon, S., Koehler-Yom, J., & Yang, Z. (2017b). The Effects of teachers' social and human capital on urban science reform initiatives: Considerations for professional development. *Teachers College Record*, *119*(4), 1–32.

Yoon, S., Klopfer, E., Anderson, E., Koehler-Yom, J., Sheldon, J., Schoenfeld, I., Wendel, D., Scheintaub, H., Oztok, M., Evans, C., & Goh, S. (2016). Designing computer-supported complex systems curricula for the Next Generation Science Standards in high school science classrooms. *Systems*, *4*(38), 1–18. Retrieved from http://www.mdpi.com/2079-8954/4/4/38/html

Yoon, S. A. (2011). Using social network graphs as visualization tools to influence peer selection decision-making strategies to access information about complex socioscientific issues. *Journal of the Learning Sciences*, *20*(4), 549–588.

Yoon, S. (2008). An Evolutionary approach to harnessing complex systems thinking in the science and technology classroom. *International Journal of Science Education*, *30*(1), 1–32.

Yuan, J., & Kim, C. (2014), Guidelines for facilitating the development of learning communities in online courses. *Journal of Computer Assisted Learning, 30*, 220–232.

Zeichner, K., & Liston, D. (2014). *Reflective teaching: An Introduction*. New York, NY: Routledge.

Ioannou, M., & Ioannou, A. (2020). Technology-enhanced Embodied Learning: Designing and Evaluating a New Classroom Experience. *Educational Technology & Society*, *23* (3), 81–94.

Technology-enhanced Embodied Learning: Designing and Evaluating a New Classroom Experience

Marianna Ioannou[1] and Andri Ioannou[1,2*]
[1]Cyprus Interaction Lab, Department of Multimedia and Graphic Arts, Cyprus University of Technology, Cyprus // [2]Research Centre on Interactive Media, Smart Systems and Emerging Technologies (RISE), Cyprus // marianna@cyprusinteractionlab.com // andri@cyprusinteractionlab.com
[*]Corresponding author

(Submitted April 22, 2020; Revised June 7, 2020; Accepted August 10, 2020)

ABSTRACT: The enactment of embodied learning in the authentic classroom introduces new challenges. The educational system has yet to develop a clear vision or learning design models that would guide the implementation of embodied learning using digital technologies and manipulatives. This study presents an example of a learning design for technology-enhanced embodied learning in an authentic classroom. Three forms of physical embodiment (direct, surrogate and augmented) are enacted using a model consisting of a single educator and rotating across learning stations. The case study takes place in a multidisciplinary lesson around historical information. In this lesson, Year 4 primary school students (i) take virtual tours among the ruins of Archaic kingdoms using mobile VR headsets, (ii) use programmable floor robots to learn about the various occupations people had back then and (iii) create storyboards based on historical information using web-based digital tools. The study evaluates the technology-enhanced embodied learning experience from the perspective of the learners. Data from 34 students demonstrate learning gains, as well as positive perceptions of the learning experience in terms of their relationship with their teammates, their sense of personal development, and the overall classroom orchestration. We conclude with lessons learnt, limitations and suggestions for future work. With this study, we aim to spark a dialogue on how technology-enhanced embodied learning can be successfully enacted in real-world classrooms, highlighting the need for more studies in the intersection of technology, design and pedagogy.

Keywords: Embodied learning, Technology integration, Technology-enhanced learning, Classroom orchestration, Learning design

1. Introduction

Research on learning design is being increasingly influenced by theories of embodied cognition. These indicate that we learn better by engaging our senses through our bodies and that the mind is not an abstract and isolated entity, but rather, it is integrated into the body's sensorimotor systems (Barsalou, 1999; Barsalou, 2008). Advocates of embodied cognition argue that cognition is both perceptuomotor and modal, in the sense that we function in the world through sensation, perception, imagination and motor action, and by engaging with cultural forms, artefacts, language and practices that have evolved to support our survival and development (e.g., Georgiou & Ioannou, 2019; Hostetter & Alibali, 2008; Johnson-Glenberg, 2018). From this perspective, knowledge is grounded on the experience(s) that we collect with our senses, mediated by our body. That is, action experience has a strong influence on the cognitive and neural processes and influences thought (Goldin-Meadow & Beilock, 2010) and memory (Ianì, 2019).

Driven by an embodied view of human cognition, educational technologists and learning scientists are striving to design learning experiences that promote multisensory processing, using technology as well as tangibles and manipulatives. A large spectrum of technologies based on novel interaction modalities—ranging from multi-touch to virtual reality—have been developed to enrich these learning experiences. These technologies aim to promote sensory engagement by offering new opportunities for physically interacting with objects and digital representations, foregrounding the role of the body in interaction and learning. Currently, virtual reality (VR) appears to be the most prominent tool in enabling embodied learning. The increase in commercial VR technologies, which are both affordable and immersive, in combination with a growing availability of virtual content, provides new possibilities to introduce VR into schools and other educational institutions (Johnson-Glenberg, 2018).

This study adopts "Instructional Embodiment Framework" (IEF) proposed by Black, Segal, Vitale, and Fadjo (2012). Based on IEF, instructional embodiment is composed of two main categories -- physical and imagined embodiment. A physical embodiment may be "direct," "surrogate," or "augmented." "Direct embodiment" is achieved when the user moves in a certain way in a physical space, while “surrogate embodiment” applies when the user gives instructions to a programme. Other researchers describe this as a “manipulated form of embodiment”, referring to the use of physical objects such as tangibles and manipulatives that allow for the learning of concepts to be directly embedded into the physical material of the object, as well as through the embodied interactions learners have by manipulating these objects (Minocha, Tudor, & Tilling, 2017; Pouw, Van Gog, & Paas, 2014; Price, Sheridan, Pontual Falcao, & Roussos, 2008).“Augmented Embodiment” refers to the use of a representational system, such as an avatar, in conjunction with an augmented feedback system (such as Microsoft’s Kinect and display system) to embed the embodied learner within an augmented representational system. In addition to the physical embodiment, an individual also embodies action and perception through imagination. Imagined embodiment is characterized as the mental simulation of physically embodied action (Black et al., 2012).

There is evidence that technology-enhanced embodied learning can have a positive impact on at least one of the three domains of learning: (a) Cognitive domain, (b) Affective domain, and (c) Psychomotor domain. A number of studies reported an increase in students’ conceptual knowledge on a variety of topics related to mathematics (e.g., Arroyo, Micciollo, Casano, Ottmar, Hulse, & Rodrigo, 2017; Alibali, & Nathan, 2012), biology (e.g., Andrade, Danish, & Maltese, 2017), chemistry (e.g., Tolentino, Birchfield, Megowan-Romanowicz, Johnson-Glenberg, Kelliher, & Martinez, 2009), physics (e.g., Enyedy, Danish, Delacruz, & Kumar, 2012) or language learning (e.g., Hsiao, & Chen, 2016; Kosmas, Ioannou, & Zaphiris, 2018). Recent research has shown benefits in terms of knowledge retention as delay test indicated (e.g., Kuo, Hsu, Fang, & Chen, 2014; Johnson-Glenberg, Birchfield, Tolentino, & Koziupa, 2014). A number of studies yield students’ engagement with the learning process (e.g., Ibánez & Wang, 2015; Ioannou, Georgiou, Ioannou, & Johnson-Glenberg, 2019; Lindgren, Tscholl, Wang, & Johnson, 2016; Tolentino et al., 2009) as well as on students’ increase of motivation for participation in the task (e.g., Georgiou, Ioannou, & Ioannou, 2019; Hwang, Shih, Yeh, Chou, Ma, & Sommool, 2014; Yang, Chen, & Jeng, 2012). The richer the perceptual environment using multiple sensory modalities (e.g., using visuals, voiceovers, and movement) during initial learning the better the student learning, understanding and motivation (Han & Black, 2011). Yet, the enactment of embodied learning in a classroom setting nevertheless increases the complexity of the teaching and learning experience. The educational system has yet to develop a clear vision or learning design models that would guide the implementation of embodied learning using digital technologies and manipulatives. As argued by Ioannou, Ioannou, Georgiou, and Retalis (2020) and Johnson-Glenberg (2019), when it comes to the real-word application of emerging technologies and pedagogies, empirical studies at the crossroads of technology, design and pedagogy are lacking.

This is an investigation of a technology-enhanced embodied learning experience in the classroom. The enactment of embodied learning aims to help better build this embodied experience for better understanding of the historical information, for current learning or subsequent learning, which is conventionally based on verbal or textual input. The study is part of a larger design-based research project on technology integration for embodied learning which aims to inform the current educational practice in real-world classrooms. We begin by presenting an example of a learning design for technology-enhanced embodied learning in an authentic classroom. Three forms of physical embodiment (direct, surrogate, and augmented) are enacted in the classroom and an orchestration strategy is applied, informed by previous empirical work (Ioannou, Ioannou, Georgiou, & Retalis, 2020). The study goes on to the evaluation of the technology-enhanced embodied learning experience from the perspective of the learners.

The research questions of the study are:
- Did the students experience learning gains?
- What were their perceptions of the technology-enhanced embodied learning experience?

2. Learning design

2.1. Context

This study takes place in a multidisciplinary lesson around historical information. We employ a variety of activities with technologies (virtual reality, robot programming, tablets) each harnessing the power of embodied learning according to Black et al. (2012) IEF framework. The students engage in a virtual field trip, programme the Bee-bot, and create storyboards. The idea of expanding the curricular space by integrating technology in existing course units, thus allowing the development of problem solving and teamwork skills together with domain knowledge, is not new (Ioannou, Socratous, & Nikolaedou, 2018).

2.2. The Learning Station Rotation Model

Our previous work informed our learning design in this study. We adopted the Learning Station Rotation Model (Ioannou, Ioannou, Georgiou, & Retalis, 2020), which allows students to rotate through learning stations on a fixed schedule. In our own prior research, we presented designs and orchestration strategies for technology-enhanced embodied learning that took into consideration real classroom realities: a limited access to technology, a single teacher handling 15 or more students, a curriculum that needed to be covered, as well as the teachers' aim to enact constructivist and student-centered pedagogy.

In this study, the classroom was organised into three stations. The students were split into three groups of six students each (i.e., they rotated in groups of six). They worked individually (for station 1) or in pairs (for stations 2 and 3). At station 1, the students used mobile VR headsets and the Google Expeditions app. At the second station, the students used programmable floor robots (Bee-bots). At the third station, the students used tablets to create storyboards. Each station could operate independently and there was no need for the students to follow a sequential order. For the last 15 minutes of class, the students converged for a plenary discussion of major ideas from the overall experience. See Figure 1 for the Learning Station Rotation Model adapted in this study.

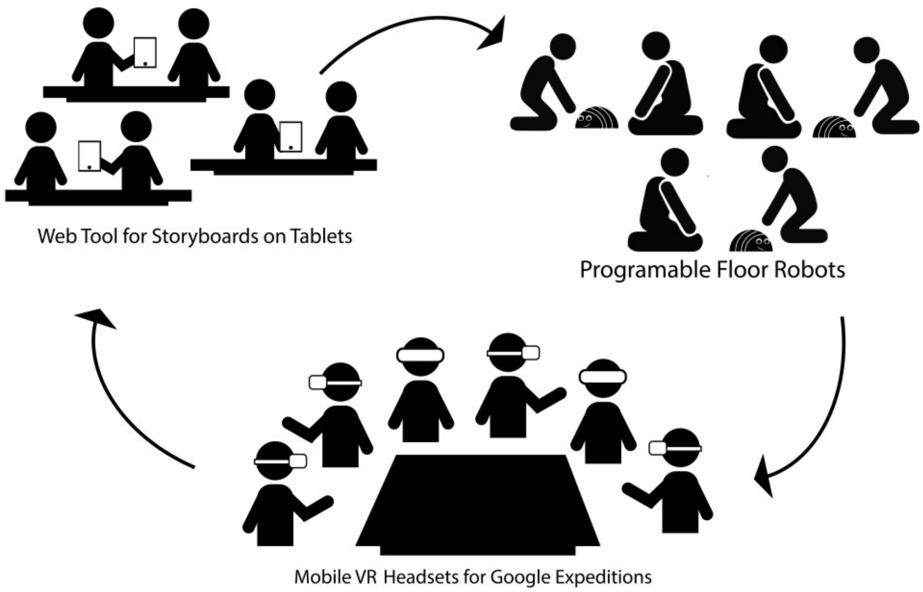


Figure 1. Learning Station Rotation Model

2.2.1. Station 1 – Mobile VR headsets and Google Expeditions

In recent years, affordable, commercial virtual reality technologies have proliferated, along with the availability of virtual content, providing new opportunities to bring VR to schools and other educational institutions (Johnson-Glenberg, 2018). VR technologies offer vivid and immersive audiovisual interfaces for eliciting bodily activity (Lindgren et al., 2016; Lindgren & Johnson-Glenberg, 2013). With VR, body-based experiences are more perceptually immersive, allowing learners to experience a more authentic and meaningful educational space (Dede, 2009).

This first station aimed to enable students to go on virtual field trips and visit three archaeological sites in their own country. The researchers developed a 360VR experience using a free tool called Google Expeditions. The VR experience is based on tours synthesized of images taken at 360 degrees (i.e., spherical images) from Kourion, Amathus, and Idalion respectively, three sites that feature remains of Archaic kingdoms.

In the VR preproduction phase, we specified the teaching and learning goals and storyboarded the guided tour for each site based on historical information. In the postproduction phase, we used a 360 camera to take the spherical images and created our own tour on the Google Expeditions platform. This was done by importing our images and creating a guided tour based on our storyboard. The Tour Creator app was another platform we used: it allowed us to add multiple scenes in the tour using the uploaded 360 images, as well as sounds, points of interest, and descriptions of each scene. Once the VR tour was ready, the students opened the Google Expeditions app on their smartphones and placed their smartphone into the mobile VR headset (The teacher guided the students through the expedition using a tablet or smartphone; s/he selected points of interest and asked questions that encouraged exploration and discovery (rather than yes/no type of questions). See Figure 3 on how Google Expeditions works and Figure 2 for screenshots from the virtual tours.

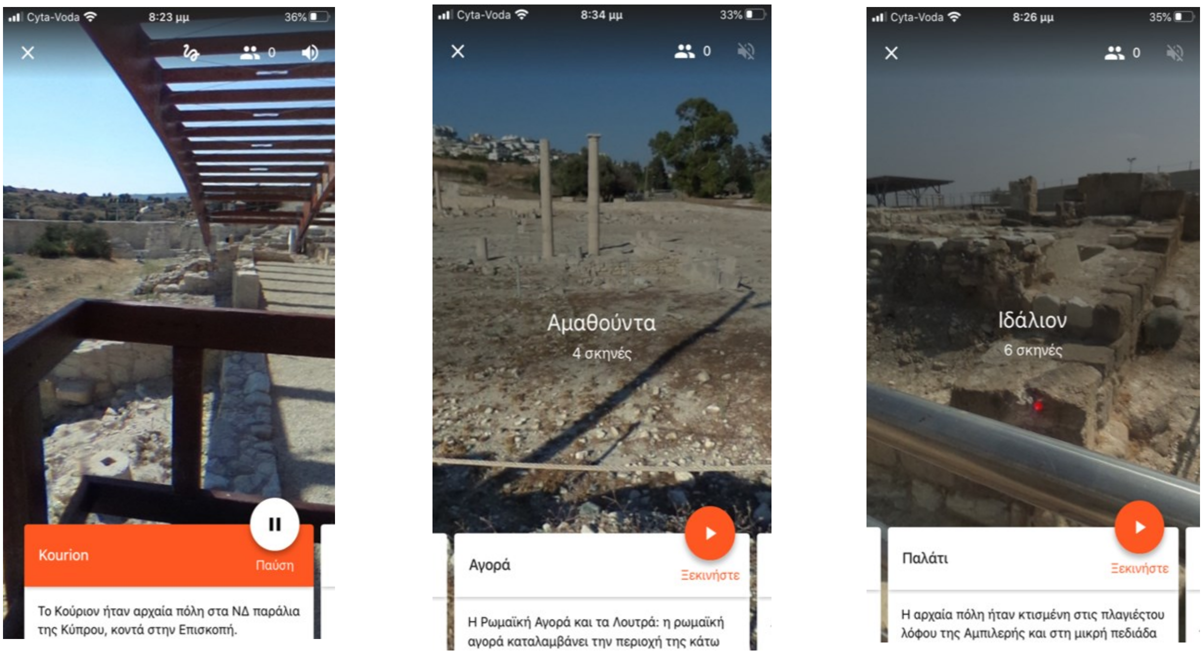


Figure 2. Google Expeditions – screenshots from virtual tours of Kourion, Amathus, and Idalion

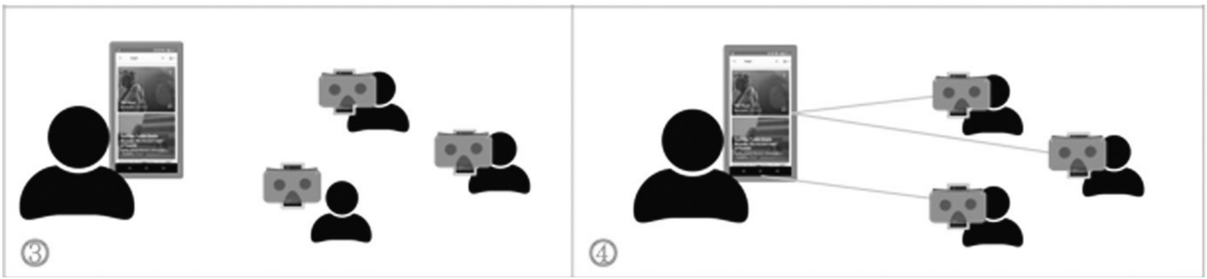


Figure 3. Google Expeditions – How it works with students as explorers and the teacher as a guide

Figure 4. Station 1- Mobile VR headsets and Google Expeditions tours

The main learning goals of the VR activity were (i) to enable students to describe similarities concerning soil morphology among the archaeological sites and (ii) to help them understand the choices made by Archaic people in order to establish a new settlement. Students used mobile phones along with a compatible mobile VR headset, which allowed them to turn and move as they would in the real world. The digital setting responded to the learner’s movements, the visuals and audio changing naturally to give a sense of reality. Being able to see evidence of the real world, even in the periphery, maintained an illusion of presence, such that learners felt their bodies were inside the virtual environment. The virtual tour was driven by the teacher who acted as a "guide" for the students in the virtual world, encouraging them to examine points of interest. Some unstructured exploration was useful in the first few minutes for students to get used to the headset and also indulge their sense of curiosity, especially for those having their first experience in a VR environment. Students at station 1 worked individually (see Figures 3 and 4).

2.2.2. Station 2 – Programmable floor robots

Physical tools such as tangibles (e.g., robots) are used in our embodied learning activities e.g., learning to programme (Black et al., 2012; Price et al., 2008). Children engage in a unique process of action and reflection that can lead to abstract thinking (Price et al., 2008). According to Black et al. (2012) IEF framework the embodiment through manipulative falls into the category of "surrogate embodiment" which is physical embodiment that is controlled by the learner whereby the manipulation of an external "surrogate" represents the individual.

In this second learning activity, students explored the occupations people had in Archaic times, while also indirectly learning to think computationally. The playful learning activity made use of programmable robots called Bee-bots. Students had to programme the toy-like Bee-bot to move on a paper mat with images representing occupations in the Archaic era. The students had to choose a description of an occupation from a set of 16 flash cards, understand the description, then programme the Bee-bot to get to the image of the respective occupation on the paper mat. Students at station 2 worked in pairs (Figure 5). Understanding the occupation was not straightforward and required discussion and agreement between the teammates.

Figure 5. Station 2- Programmable floor robots

2.2.3. Station 3 – Using tablets to create storyboards

Touching the objects on a screen directly with fingers, rather than having a control device such as a mouse or a stylus, can enhance the haptic channel experience and make the learning experience more relevant to the learning content (Black et al., 2012). Gestural interfaces (also known as natural user interfaces) include touch interfaces and free-form interfaces. Touch use interfaces (TUIs) require the user to touch the device directly and are based on a single or multi-touch point. Free-form gestural interfaces do not require the user to touch or handle the device directly (e.g., Microsoft Kinect) (Black et al., 2012). Gestural interfaces suggest new opportunities to include touch and physical movement that can benefit learning, in contrast to the less direct, somewhat passive mode of interaction suggested by using a mouse and keyboard.

The learning goal at station 3 was for the students to learn about myths of the Archaic period through the creation of guided storyboards. A set of how-to sheets helped the students explore the myths of their country's Archaic period. The software utilised included easily customisable templates and a variety of characters and backgrounds ideal for this activity. Students could also create dialogues and narratives for their stories. Students were familiar with the use of the storyboarding tool from previous class activities. Students at station 3 worked in pairs (Figure 6).

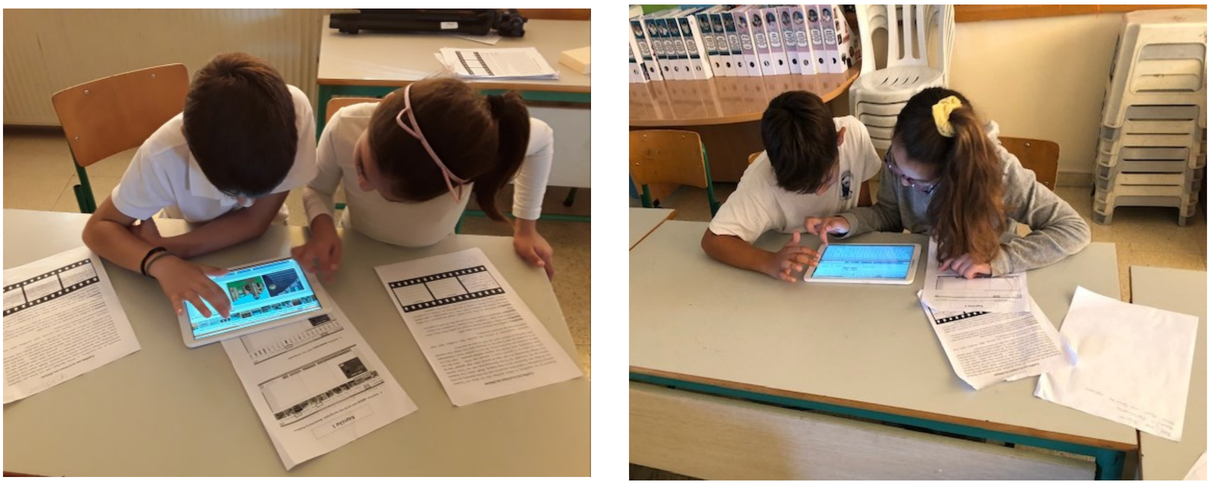


Figure 6. Station 3- Using tablets to create storyboards

3. Methodology

3.1. Research design

This study is part of a broader design-based project investigating the design of technology-enhanced learning settings grounded on embodied cognition theories, which involve orchestration of learning activities in a real classroom setting. The case study uses a mixed method approach to data collection and analysis. According to Creswell and Plano Clark (2011), collecting, analysing and mixing both quantitative and qualitative data in a single study can provide a better understanding of the research question under investigation.

3.2. Participants

The participants were Year 4 students (N = 34, 8-9-year-olds) from two classes at a public primary school in the Eastern Mediterranean. 16 participants came from one class (7 girls, 9 boys) and 18 from the other (10 girls, 8 boys). All provided parental consent. The students had not previously used Google Expeditions; it was also their first-time using VR headsets at school. The students were however familiar with the technology used at stations 2 and 3 (the Bee-bots and the storyboarding software).

3.3. The learning intervention

The learning interventions lasted 80 minutes in each class. Students were divided into three groups of six students each. The students worked alone at station 1 and in mixed-ability pairs at stations 2 and 3. Mixed-ability pairs were formed based on their teachers' knowledge of students' academic background and learning needs, collaboration skills and social relationship. The intervention started with the teacher briefly presenting the classroom setup and the task for each station. Each student group then had 20 minutes to work at each learning station. At station 1, the students went on Google Expeditions tours using their mobile VR headsets. At station 2, the students used the programmable floor robots (Bee-bots). At station 3, the students used tablets to create storyboards. See Figures 4-6 for a snapshot of station activities. The students transitioned from station to station at the sound of a bell. The intervention concluded with a 10-minute debriefing of the learning activities, where the teacher asked the students to reflect on their findings and understanding of the activity. Data collection was done immediately prior and following the intervention.

4. Date collection and analysis

The study adopted a mixed method design, using quantitative and qualitative data.

4.1. Understanding of historical information (pre-post-test)

We used a pre-post-test to assess the students' knowledge gains from the experience. The test assessed their understanding of the historical period based on three short-answer knowledge questions (KQs). The three questions were:

(KQ1) You rule a kingdom during the Archaic era. You have decided to rebuild your kingdom in a new area. Where would you choose to build it and why?
(Possible answers: (i) on a plain, (ii) near the sea for trade, (iii) near the river, (iv) in an area with fertile soil, (v) in an area that is a natural fortress, (vi) near an area with mines);

(KQ2) How and by whom do you think the new kingdoms were founded during the Archaic years?
(Possible answers: (i) by people moving within Cyprus and looking for a better place to live, (ii) by people who came from neighbouring areas such as Greek settlers, heroes of the Trojan War, Phoenicians or other neighbouring peoples);

(KQ3) If you were a resident of an Archaic kingdom, what professional choices do you think you would have? (Students could name up to 10 professions).

Multiple answers were possible, and points were awarded for historical accuracy. The total score was awarded out of 100: 5 points for each answer to the first question (maximum of 30 points), 10 points for each answer to the second question (maximum of 20) and 5 points for each answer to the third question with a maximum of 50 points. A paired sample t-test was conducted to assess learning gains from pre to post testing.

4.2. Perceptions of technology integration (post-test)

A post-interventional questionnaire evaluated the students' perceptions of technology. The questionnaire was slightly modified from previous studies conducted by Ioannou, Ioannou, Georgiou, and Johnson-Glenberg (2019) and other researchers (Wu, Chang, & Guo, 2009; Maor & Fraser, 2005), and presented evidence of reasonable internal consistently in the present study. The questionnaire assessed technology integration on three subscales, using a 5-point Likert agreement:

- "Relationship" (9-items scale's Cronbach's alpha = .705), assessing the extent to which students had opportunities to discuss their ideas and support each other (e.g., I asked other students to explain their ideas; students were willing to help each other).
- "Personal development" (10-items scale's Cronbach's alpha = .615), assessing the extent to which students were motivated to learn and think about their personal learning (e.g., students set up study goals on their own; I got to think deeply about what I was learning).
- "System maintenance and change" (5-items scale's Cronbach's alpha = .755), assessing the extent to which the "system" was easy to work with (e.g., it was easy to learn how to use the stations; the setup was fun).
- The data collected were analysed via descriptive statistics.

4.3. Focus groups

When the experience was completed, we conducted two focus groups, each with eight students selected by the class teachers to create a mixed group in terms of gender and ability. Each focus group lasted 25 minutes and aimed to understand the students' overall learning experience. Driven by Moos' (1987) conceptual framework of technology integration, the students were probed to discuss their experience in terms of: (a) their personal development (i.e., what were the main factors that helped you learn while participating in this experience?), (b) their relationships with others (i.e., how did you collaborate with team members at each learning station?), and (c) system maintenance and change (i.e., did you encounter any problems while using the technologies in the learning stations? How did those problems affect you?). The interviews were all transcribed verbatim. A thematic analysis was conducted focusing on the aforementioned three dimensions of our conceptual framework: "Relationship," "Personal development" and "System maintenance and change." The analysis was done by two independent coders with nearly 90% agreement; all disagreements were resolved following a discussion between the coders.

5. Findings

5.1. Understanding of historical information (pre-post-test)

A paired-samples t-test of pre-post total mean scores across the three knowledge questions, indicated significant learning gains from pre (M = 0.85, SD = 0.56) to post testing (M = 1.65, SD = 0.56), $t(34) = -7.05$, $p < .001$, with a large effect size (d = 1.43). Table 1 present the paired-samples t-test per individual knowledge question, showing statistically significant gains in all three questions. Overall, the embodied and immersive learning experience as enacted in this work, appears to have enabled students' learning and understanding about the historical information under consideration.

Table 1. Paired-Samples *t*-test per Knowledge Question

	Mean (post test - pre test)	Std. deviation	*t*	*df*	*Sig.* (2-tailed)
KQ1	.53	.90	-3.45	33	.002**
KQ2	.35	.85	-2.43	33	.021*
KQ3	2.50	1.24	-11.79	33	.000***

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

5.2. Perceptions of technology integration (post-test)

Descriptive statistics demonstrated positive attitudes on all three dimensions of the technology integration questionnaire. The mean scores were well above the midpoint of the scale (the median of the scale is 3 and total score is 5), namely: (i) M = 4.33 (SD = 0.47) for "Relationships", suggesting that the students discussed their ideas and were supportive to each other; (ii) M = 3.60 (SD = 0.59) for "Personal development," suggesting that the students were motivated to learn and thought about their own learning; and (iii) M = 4.53 (SD = 0.50) for "System maintenance and change," suggesting that the learning stations, as a system, were easy and fun to use.

5.3. Focus groups

What follows is an overview of the core ideas discussed by the students, coded and organised within the three dimensions of our conceptual framework.

5.3.1. Relationships

Students reported on the value of teamwork for their learning, especially with reference to the programmable robot station and the storyboarding station. More specifically, the students spoke about the exchange of views and ideas that took place at these two stations, as well as peer assistance and collaboration.

> I liked working in pairs because we could play for longer, agree on how to do our assignment, and help each other. [Boy5 - focus group 1]
>
> I liked that my teammate L. was sharing her thoughts with me on how we could improve our storyboard. [Girl6 - focus group 2]
>
> I had some difficulties figuring out how to do the storyboard, but P. showed me how because she could do it easily. [Girl5 - focus group 2]
>
> I worked very well with my teammate. We shared our thoughts and helped each other. For example, I was reading my flash card and we were both trying to guess the occupation mentioned. [Girl4 - focus group 2]

However, students reported negatively on VR's lack of multiplayer functionality.

> It would have been better to work as a group in VR. In the virtual tour I found myself trying to find where the other members of my team were, I kept asking them, "where are you?" and "what do you see around you?" [Boy6 - focus group 1]

5.3.2. Personal development

Students reported that the activity had a positive impact on their personal learning development, in the sense that it motivated them to learn and think about their personal learning.

> VR was a great way to learn about where Archaic kingdoms where built. At the other two stations, I was motivated to learn about people's occupations and about the myths of the Archaic era. [Girl1- focus group 1]

> I liked using VR because I was seeing everything right there… and I liked the feeling of being there and watching it like it was real. It was fun the learn this way. [Girl2 - focus group 1]

> I liked the Bee-bot station because I had to think about how to programme the Bee-bot robot to move in a certain way. We had to read the flash card first and correctly guess each occupation. I liked the other stations too… it was a creative exercise, doing our own storyboard and writing our own dialogues based on myths we learned in our History lessons. [Girl4 - focus group 2]

> The arrows showed where I had to look while walking around. The teacher asking questions helped me think about the soil morphology and take better notice of the points of interest. [Girl3 - focus group 1]

The students also commented on how the experience was fun and engaging; they preferred it to a typical learning experience.

> It was also fun and easy learning. [Girl4 - focus group 1]

> I liked the creation of a storyboard because I could use my creativity. [Girl3 - focus group 2]

> Using VR, I felt like I was taking a school trip to an archaeological site. It was so much more fun than a typical lesson or even actually going on a field trip. [Girl2 - focus group 1]

Students did, however, report two disengaging factors that inhibited their personal development: time pressure and a lack of feedback at stations 2 and 3.

> We need more time at the stations. For example, at station 3 with the storyboard, we only had five minutes to write down the dialogue, and it was difficult to finish on time. [Boy1 - focus group 2]

> Sometimes we wanted to ask our teacher a question, but she was busy at the VR station and we had to wait until we got her attention, or we had to ask our teammates to help us. We were wasting time waiting for help. [Boy8 - focus group 2]

5.3.3. System maintenance and change

The students were positive about the "system": they reported it was easy and fun to work with. They further commented that VR allowed for a feeling of presence and this had a positive impact on their learning. Participants reported that they were surprised by how immersed they felt within the VR sites.

> It was a fun and easy way to learn History. It was like a typical lesson but enriched with technology and games, which made it fun to learn about historical content. [Boy2 - focus group 2]

> The virtual reality activity was amazing because we could see the places and feel like they were real and you were really there at the archaeological site. [Boy5 - focus group 1]

Students reported negatively on two aspects of the exercise: the discomfort caused by the size of the VR headset and also the inconvenience of having to sit on the floor while working with the programmable floor robots.

> I liked the virtual tour, but it felt annoying here (pointing to his face) because the headset was loose and I had to hold it to my face when I was looking down. The teacher tried to adjust the size but it was still loose. [Boy7 - focus group 1]

> I liked that I could work with my teammate in our own space, but it was difficult to work on the floor with the Bee-bot. It would have been better if we had some cushions to sit on or had sat at our desks. [Boy8 - focus group 2]

6. Discussion and implications

This study aimed to investigate a technology-enhanced embodied learning experience in multidisciplinary lesson around historical information. We presented an example of a learning design enacting three forms of physical embodiment (direct, surrogate, and augmented) in the classroom, using a model of rotating across learning stations, orchestrated by a single educator. Results from 34 learners demonstrated learning gains, as well as positive perceptions of the learning experiences in terms of "Relationship," "Personal development" and "System maintenance and change." As part of a larger design-based research project, this study aimed to inform the current educational practice on technology-enhanced embodied learning in real-world classrooms. Below, we reflect further on our findings.

With respect to RQ1 (Did the students experience learning gains?), responses to the knowledge test, as well as reporting in the focus groups, revealed that the technology-enhanced embodied learning approach managed to transform the experience of discovering new places, understanding spatial relations and learning historical facts, making it both enjoyable and effective. In the focus groups, the students said the virtual field trips were an effective and intriguing way of learning. They appreciated the teacher's guidance through the VR experience which, they said, helped them learn more effectively. This finding is consistent with other authors writing about VR in education, who have argued that embodied learning can enhance involvement in learning processes (e.g., Chittaro & Buttussi, 2015; Georgiou, Tsivitanidou, Eckhardt, & Ioannou, 2020; Jha, Price, & Motion, 2020; Skulmowski & Rey, 2018). It appears that the design of the learning stations allowed students to engage with a variety of information and gain knowledge in a fun way.

With respect to RQ2 (What were their perceptions of the technology-enhanced embodied learning experience?), the students seem to have had an overwhelmingly positive learning experience in terms of "Relationship," "Personal development" and "System maintenance and change," as evident in both the quantitative and qualitative data. The Learning Station Rotation Model seems to have a significant impact on students' engagement. They were active, managed themselves, and solved problems in the context of each learning station. The students elaborated that the technology- and manipulatives- enhanced stations were preferable to conventional ways of classroom learning and that they constituted an attractive and fun learning environment that fueled their interest and curiosity. This finding is in line with Minocha et al. (2017) who argued that, because the students are in control of where they look and for how long, they can follow their interest and curiosity, hence giving them a sense of empowerment over their own exploration. Johnson-Glenberg (2018) also found that whenever users felt they had control over the environment, they experienced agency, which is in line with the reporting of the students in the present study.

Overall, the effectiveness of the VR field trips was related to the concept of presence and immersion. Presence refers to users' subjective belief that they are in a certain place, even if they know that the experience is mediated by the computer (Schuemie, Van Der Straaten, Krijn, & Van Der Mast, 2001; Slater, 1999). In VR heritage scenarios, "cultural presence" plays a key role; it's not just a feeling of "being there" but of being - not only physically, but also socially, culturally - "there and then" (Champion, 2010). Although the low-cost development in this work did not offer the students a highly interactive experience with virtual objects - which would have been possible with advanced VR tools like Oculus Rift or HTC Vive - the experience was perceived as immersive in terms of presence and thus highly enjoyable. The enactment of embodied learning aimed to help better build this embodied experience for better understanding of the historical information, for current learning or subsequent learning, which is conventionally based on verbal or textual input. Indeed, compared to a conventional learning design, learners in an embodied experience can get immersed in the virtual context and engage with the learning content, getting realistic information on abstract and complicated concepts or artefacts. The VR field trip was designed to offer a virtual but authentic learning context in which learners could imagine what a real field trip to a site would be like. This immersive and interactive experience from the comfort of the classroom gave them meaningful learning moments without the expense of long journey. Today's affordable motion-sensing input devices, together with freely available apps such as Google Expeditions, can support learning design for embodied interaction and could provide solutions when mobility is costly and not always possible. With this study, we aim to encourage more educators to take advantage of these affordable tools.

In this study we used affordable learning technologies enriched with content that is aligned with the national educational curriculum. We aimed to combine the physical and the digital worlds and to enable a multisensory and embodied learning experience to promote an understanding of historical information in a multidisciplinary lesson. Our focus was on low-cost technologies as schools often lack the financial resources, technological infrastructure and professional development for teachers. Therefore, we recommend the deployment of low-cost and easily built VR environments (Kalpakis, Palaigeorgiou, & Kasvikis, 2018; Palaigeorgiou, Karakostas, &

Skenteridou, 2018), as well as apps that can be integrated with existing educational curricula or are flexible enough for teachers to edit the content based on students' needs and learning goals (Ioannou, 2018). With this study, we aim to encourage more educators and learning designers to develop content and share their experiences.

The students had a negative response to the VR's single-player mode: they expressed a desire to meet up in the virtual space, suggesting that the experience was indeed immersive and that they wanted to be at the same place, at the same time. This could confirm the previously reported human need for social exploration of heritage sites using a mobile phone guide (Suh, Shin, & Woo, 2009). Learning in the classroom, especially for this age range of students, is a fundamentally social activity but most VR technologies, like the one used in this study, do not currently offer a group mode for collaborative learning. Therefore, VR embodied learning experiences should ideally unfold within a well-structured group learning context or follow a scenario with embedded teamwork. This may include individual work and teamwork, along with class-wide activities or plenary discussions.

Moreover, the lack of feedback from the teacher at stations 2 and 3 was commented on by the students as a negative factor in terms of their personal development. During the design of the experience, the thinking was that thanks to the tangible interface of the technologies used at stations 2 and 3, the teacher would easily be able to monitor the progress of each group while remaining at the first station. However, it turned out to be difficult for the teacher to manage the VR station guided tour while also keeping track of the progress of the groups at the other two stations. While guiding the virtual tour, the teacher could only keep a visual track of whether the students in that group were progressing with the activities. The station rotation model (Ioannou, Ioannou, Georgiou, & Retalis, 2020) would have worked better, if the virtual field trip was self-guided rather than guided by a teacher, or if some form of collaboration between teammates was in place (e.g., one student guiding another student based on how-to sheets). This scenario warrants future investigation and might also address the negative comments students had about the VR single-player mode.

Some more negative comments concerned the headsets and the classroom setup. The physical features of VR headsets play a role in the overall user experience. In this study, the subjects were pre-teen children and the mobile VR headset was too big for some (even after fastening the strap as tightly as possible). This caused discomfort for some students as they had to hold the headset up to their faces. The students also reported that it was tiresome to work on the floor for station 2; they would have preferred to have cushion or work at their desks. Future studies would do well to address these issues for young learners.

In closing, we conclude that the students' performance and input revealed positive learning gains and attitudes and underlined that the learning station model, the use of technologies and manipulatives, and the design of the learning activities were successful in providing an engaging embodied learning experience. The negative feedback from the focus groups concerning the VR's single-player mode will guide the next stage of this design-based research. First, in future work we aim to capitalise on the design for collaboration around embodied learning technologies that do not naturally encourage collaboration (e.g., VR single-user mode). Collaborative embodiment in technology-enhanced interventions has not been discussed in the literature; relevant work will help to address the development of pedagogical strategies that involve groups of students in technology-enhanced embodied learning. Finally, in the next study, we would aim to investigate the long-term retention of knowledge from this experience, ideally in comparison to some control group receiving traditional teaching on the same lesson. Closing, this study mostly relied on self- reported and retrospective measures. Future study could use in situ measurements (e.g., observation protocols, field notes and log files) for investigating how the learning process unfolds in a technology-enhanced embodied learning intervention such as the one presented here.

7. Conclusion

This study aimed to investigate a technology-enhanced embodied learning experience in a multidisciplinary lesson around historical information. We presented an example of a learning design for technology-enhanced embodied learning in an authentic classroom. Data from 34 students demonstrated learning gains as well as positive perceptions of the learning experiences in terms of "Relationship," "Personal development" and "System maintenance and change." With this study, we aim to spark a dialogue on the successful enactment of technology-enhanced embodied learning in the classroom, highlighting the need to consider more studies at the intersection of technology, design and pedagogy. Technologies enabling embodied interactions continue to offer an enormous range of opportunities and thus deserve major consideration and investigation on how they can be applied in the context of mainstream classrooms.

Acknowledgement

This work is part of the project that has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 739578 and the government of the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development.

References

Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *Journal of the Learning Sciences*, *21*(2), 247-286.

Andrade, A., Danish, J. A., & Maltese, A. V. (2017). A Measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, *4*(3), 18-46.

Arroyo, I., Micciollo, M., Casano, J., Ottmar, E., Hulse, T., & Rodrigo, M. M. (2017). Wearable learning: multiplayer embodied games for math. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (pp. 205-216). doi:10.1145/3116595.3116637

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and brain sciences*, *22*(4), 577-660.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617-645.

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

Champion, E. (2010). Playing with the past. In *Playing with the Past. Human-Computer Interaction Series* (pp. 129-155). London, UK: Springer.

Chittaro, L., & Buttussi, F. (2015). Assessing knowledge retention of an immersive serious game vs. a traditional education method in aviation safety. *IEEE transactions on visualization and computer graphics*, *21*(4), 529-538.

Creswell, J. W., & Plano Clark, V. L. (2011). Choosing a mixed methods design. *Designing and conducting mixed methods research*, *2*, 53-106.

Dede, C. (2009). Immersive interfaces for engagement and learning. *Science*, *323*(5910), 66-69.

Enyedy, N., Danish, J. A., Delacruz, G., & Kumar, M. (2012). Learning physics through play in an augmented reality environment. *International journal of computer-supported collaborative learning*, *7*(3), 347-378.

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

Georgiou Y., Ioannou A., & Ioannou M. (2019). Investigating children's immersion in a high-embodied versus low-embodied digital learning game in an authentic educational setting. In D. Beck et al. (Eds.), *Immersive Learning Research Network. iLRN 2019. Communications in Computer and Information Science* (Vol 1044, pp. 222-223). doi:10.1007/978-3-030-23089-0_17

Georgiou, Y., Tsivitanidou, O., Eckhardt, C., & Ioannou, A. (2020). Work-in-progress—A Learning experience design for immersive virtual reality in physics classrooms. In *2020 6th International Conference of the Immersive Learning Research Network (iLRN)* (pp. 263-266). IEEE. doi:10.23919/iLRN47897.2020.9155097

Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: The Case of gesture. *Perspectives on psychological science*, *5*(6), 664-674.

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

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495-514.

Hsiao, H. S., & Chen, J. C. (2016). Using a gesture interactive game-based learning approach to improve preschool children's learning performance and motor skills. *Computers & Education*, *95*, 151-162.

Hwang, W. Y., Shih, T. K., Yeh, S. C., Chou, K. C., Ma, Z. H., & Sommool, W. (2014). Recognition-based physical response to facilitate EFL learning. *Educational Technology & Society*, *17*(4), 432-445.

Ianì, F. (2019). Embodied memories: Reviewing the role of the body in memory processes. *Psychonomic bulletin & review, 26*(6), 1747-1766.

Ibánez, J. D. J. L. G., & Wang, A. I. (2015). Learning recycling from playing a Kinect game. *International Journal of Game-Based Learning (IJGBL)*, *5*(3), 25-44.

Ioannou, A. (2018). A Model of gameful design for learning using interactive tabletops: Enactment and evaluation in the socio-emotional education classroom. *Educational Technology Research and Development, 67*(2), 277-302.

Ioannou, A., Socratous, C., & Nikolaedou, E. (2018). Expanding the curricular space with educational robotics: A Creative course on road safety. In *European Conference on Technology Enhanced Learning - ECTEL 2018* (pp. 537-547). doi:10.1007/978-3-319-98572-5_42

Ioannou, M., Georgiou, Y., Ioannou, A., & Johnson-Glenberg, M. (2019). On the understanding of students' learning and perceptions of technology integration in low- and high-embodied group learning. In K. Lund, G. P. Niccolai, E. Lavoué, C. Hmelo-Silver, G. Gweon, & M. Baker (Eds.), *A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings, 13th International Conference on Computer Supported Collaborative Learning (CSCL)* (Vol. 1, pp. 304-311). Lyon, France: International Society of the Learning Sciences.

Ioannou, M., Ioannou, A., Georgiou, Y., & Retalis, S. (2020). Designing and Orchestrating the Classroom Experience for Technology-Enhanced Embodied Learning. In M. Gresalfi & S. I., Horn (Eds.), *The Interdisciplinarity of the Learning Sciences, 14th International Conference of the Learning Sciences (ICLS)* (Vol. 2, pp. 1079-1086). Nashville, Tennessee: International Society of the Learning Sciences.

Jha, R. T., Price, S., & Motion, A. (2020). Designing 'embodied' science learning experiences for young children. In *International Conference on Human-Computer Interaction* (pp. 207-225). doi:10.1007/978-3-030-50513-4_16

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

Johnson-Glenberg M. C. (2019). The Necessary Nine: Design Principles for Embodied VR and Active Stem Education. In P. Díaz, A. Ioannou, K. Bhagat, & J. Spector (Eds.), *Learning in a Digital World. Smart Computing and Intelligence* (pp. 83-112). Singapore: Springer.

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

Kalpakis, S., Palaigeorgiou, G., & Kasvikis, K. (2018). Promoting historical thinking in schools through low fidelity, low-cost, easily reproduceable, tangible and embodied interactions. *International Journal of Emerging Technologies in Learning (iJET)*, *13*(12), 67-82.

Kosmas, P., Ioannou, A., & Zaphiris, P. (2018). Implementing embodied learning in the classroom: Effects on children's memory and language skills. *Educational Media International, 56*(1), 59-74.

Kuo, F. R., Hsu, C. C., Fang, W. C., & Chen, N. S. (2014). The Effects of Embodiment-based TPR approach on student English vocabulary learning achievement, retention and acceptance. *Journal of King Saud University-Computer and Information Sciences*, *26*(1), 63-70.

Lindgren, R., & Johnson-Glenberg, M. (2013). Emboldened by embodiment: Six precepts for research on embodied learning and mixed reality. *Educational Researcher*, *42*(8), 445-452.

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

Maor, D., & Fraser, B. J. (2005). An Online questionnaire for evaluating students' and teachers' perceptions of constructivist multimedia learning environments. *Research in Science Education*, *35*(2-3), 221-244.

Minocha, S., Tudor, A. D., & Tilling, S. (2017). Affordances of mobile virtual reality and their role in learning and teaching. In *Proceedings of British HCI 2017 – Digital Make-Believe* (pp. 1-10). Sunderland, UK: BCS Learning and Development. doi:10.14236/ewic/HCI2017.44

Moos, R. H. (1987). *The Social climate scales: A User's guide*. Palo Alto, CA: Psychologists Press.

Palaigeorgiou, G., Karakostas, A., & Skenteridou, K. (2018). Touching and traveling on 3D augmented tangible maps for learning geography: The FingerTrips approach. *Interactive Technology and Smart Education*, *15*(3), 279-290.

Pouw, W. T., Van Gog, T., & Paas, F. (2014). An Embedded and embodied cognition review of instructional manipulatives. *Educational Psychology Review*, *26*(1), 51-72.

Price, S., Sheridan, J., Pontual Falcao, T., & Roussos, G. (2008). Towards a framework for investigating tangible environments for learning. *International Journal of Arts and Technology*, *1*(3/4), 351-368.

Schuemie, M. J., Van Der Straaten, P., Krijn, M., & Van Der Mast, C. A. (2001). Research on presence in virtual reality: A Survey. *CyberPsychology & Behavior*, *4*(2), 183-201.

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

Slater, M. (1999). Measuring presence: A Response to the Witmer and Singer presence questionnaire. *Presence*, *8*(5), 560-565.

Suh, Y., Shin, C., & Woo, W. (2009). A Mobile phone guide: Spatial, personal, and social experience for cultural heritage. *IEEE Transactions on Consumer Electronics*, *55*(4), 2356-2364.

Tolentino, L., Birchfield, D., Megowan-Romanowicz, C., Johnson-Glenberg, M. C., Kelliher, A., & Martinez, C. (2009). Teaching and learning in the mixed-reality science classroom. *Journal of Science Education and Technology*, *18*(6), 501-517.

Wu, W., Chang, H. P., & Guo, C. J. (2009). The Development of an instrument for a technology-integrated science learning environment. *International Journal of Science and Mathematics Education*, *7*(1), 207. doi:10.1007/s10763-007-9116-5

Yang, J. C., Chen, C. H., & Jeng, M. C. (2010). Integrating video-capture virtual reality technology into a physically interactive learning environment for English learning. *Computers & Education*, *55*(3), 1346-1356.

Tlili, A., Lin, V., Chen, N.-S., Huang, R., & Kinshuk. (2020). A Systematic Review on Robot-Assisted Special Education from the Activity Theory Perspective. *Educational Technology & Society*, *23* (3), 95–109.

A Systematic Review on Robot-Assisted Special Education from the Activity Theory Perspective

Ahmed Tlili[1], Vivien Lin[2,3,4*], Nian-Shing Chen[2,3], Ronghuai Huang[1] and Kinshuk[5]
[1]Smart Learning Institute of Beijing Normal University, Beijing, China // [2]Department of Applied Foreign Languages, National Yunlin University of Science and Technology, Yunlin, Taiwan // [3]Research Center for Smart Learning, National Yunlin University of Science and Technology, Yunlin, Taiwan // [4]Bachelor Program in Interdisciplinary Studies, College of Future, National Yunlin University of Science and Technology, Yunlin, Taiwan // [5]University of North Texas, Denton, TX, USA // ahmed.tlili23@yahoo.com // vivienster@gmail.com // nianshing@gmail.com // huangrh@bnu.edu.cn // kinshuk@ieee.org
[*]Corresponding author

(Submitted March 18, 2020; Revised May 30, 2020; Accepted August 10, 2020)

ABSTRACT: The design, implementation, and outcome of educational robots in special education have not been sufficiently examined in a systematic way. In particular, learner-based and contextual factors, as well as the essential roles played by various stakeholders have not been addressed when robots are used as a learning tool in special education. Therefore, a systematic review using Activity Theory was conducted to analyze 30 studies in robot-assisted special education. Content analysis of the studies reported relevant information with respect to each activity component ─ (a) *subject* (learners with disabilities), (b) *technology* (robots supported by instructional design), (c) *object* (target skills or behaviors), (d) *rules* (implementation procedure and performance measures), (e) *community* (learners with disabilities, special education professionals, and parents), (f) *division of labor* (among learners, professionals and parents), and (g) *outcome* (performance of target skills or behaviors). Furthermore, the study identified existing gaps from the robot-assisted special education studies (e.g., lack of parental engagement), challenges (e.g., difficulty with standardizing performance measures due to heterogeneity of learner profiles), and contradictions (e.g., opposing views among experts on the role of robots in social interactions). Finally, recommendations were made under each activity component. The study concluded that both general and domain-specific guidelines should be created for each disability category proposed in this review to assist practitioners who wish to use robots to assist special education.

Keywords: Educational robots, Special education, Disability, Activity Theory, Human-robot interaction, Assistive technology

1. Introduction

In line with the re-authorization of the Individuals with Disabilities Education Act (108^{th} Congress, 2004), special education professionals face the urgency to identify and implement effective practices to ensure the benefits of learners with disabilities (Moeller et al., 2015). The Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan (n.d.) defines special education as education that aims to take into consideration the individual needs of learners with disabilities in order to develop both their social participation and independence. In recent decades, special education practitioners are making use of assistive technologies, which are services and technology devices that facilitate the learning process among individuals with disabilities (Ashton, 2005). Robots, as an example of assistive technologies, can effectively help learners with disabilities fulfill the requirements of subject learning (e.g., science, art) and play activities when supported by appropriate strategies (Encarnaçao et al., 2016). Although many studies have applied educational robots in special education settings, a systematic analysis on the related design, implementation, and outcome using robots in special education interventions has not been conducted, nor have applicable guidelines for robot-assisted special education been established. This review therefore aims to fill this gap.

Among various types of educational technology, robots are considered a safe and accessible learning tool to support children's flexible and programmable manipulation of real objects (Encarnação et al., 2016). The fun and engaging learning experience also lead to better learning achievements (e.g., Datteri et al., 2013). In particular, the rule-based, programmed robot-human interactions effectively keep learners with Autism Spectrum Disorder (ASD) engaged in learning. Moreover, for learners with disabilities, robots can increase their readiness to learn in the classroom through predictable and consistent human-robot interactions, which makes robots an important transition tool for helping ASD learners to progress from human-robot to human-human interactions (Alcorn et al., 2019). As current literature only provides a partial picture about the integration of robots in education, a comprehensive investigation on how to deploy robots with effective design and strategies to enhance learning in special education settings is needed. Consequently, the scope of this review includes how various types of robots have been used in combination with instructional design strategies and implementation processes in robot-assisted special education.

To understand and optimize the process of robot-assisted special education, a systematic examination of its key components is needed. This review study first applies the Activity Theory (Engeström, 2001) to analyze the contribution and agency of the stakeholders involved in the practice of robot-assisted special education, and then proposes recommendations on how to best integrate educational robots as a learning technology (Cheng, Sun, & Chen, 2018) for learners with disabilities. The Activity Theory describes the interconnectedness among social and individual processes in an activity supported by a mediating tool. Six main components of the theory include *subject*, *object*, *tool*, *rules*, *community*, and *division of labor* (Engeström, 2001). Since complex and interacting factors affect the design and perceptions of tools in special education (Pearson, 2009), it is crucial to move beyond the technology itself (robots in this case) and understand how different stakeholders can best collaborate (e.g., identifying the disability profiles, standard practices, and distributed duties) to achieve a goal. As such, an analysis based on the Activity Theory, which has been applied in numerous domain areas, will help to increase the effectiveness of using robots in special education settings.

Activity Theory has been adopted to examine the use of several types of technology in both general and special education settings (Daniels & Cole, 2002; Edwards et al., 2002; Pearson & Ralph, 2007). However, the theory has not been used to analyze the design and implementation of robot-assisted special education. In a systematic review on robotics education, Jung and Won (2018) suggested that robotics education research should shift its focus from the effects of robotic technology to learners by examining the pedagogies, teaching methods, and specific ways learners undergo meaningful learning processes. As reviews on learner-centered design and the implementation of robotic assistive technology in special education remain scant (e.g., Van den Heuvel et al., 2016), this study probes into how various components work in a special education activity system that employs robots by asking two research questions:

RQ1: What relevant features concerning the design, implementation, and outcome of robot-assisted special education can be identified through the lens of Activity Theory?
RQ2: What recommendations can be made to improve robot-assisted special education research?

2. Method

This systematic review aimed to analyze existing practices in designing, implementing, and measuring outcome in robot-assisted special education. The researchers identified the need for a review with specific research questions, selected the studies for review, assessed their quality, and presented the data extraction method with interpretations and recommendations for further research (Benitti, 2012; Kitchenham, 2004).

2.1. Search strategy and inclusion/exclusion process

Systematic searches included nine entries with keywords such as "educational robots" AND "learners with disabilities" on eight electronic databases. Figure 1 provides the flow chart of the entire selection process. The initial number of studies after the keyword searches was 784 after removal of duplicates. Then, two of the researchers conducted title, abstract, and full-text screening using ten inclusion/exclusion criteria with good inter-coding reliability ($\kappa = 0.83$). Consensus was reached via discussion for any discrepancy in the selection results. This led to a final data set of 30 studies, included twenty-seven journal papers (6 SSCI journals), two conference papers, and one book chapter.

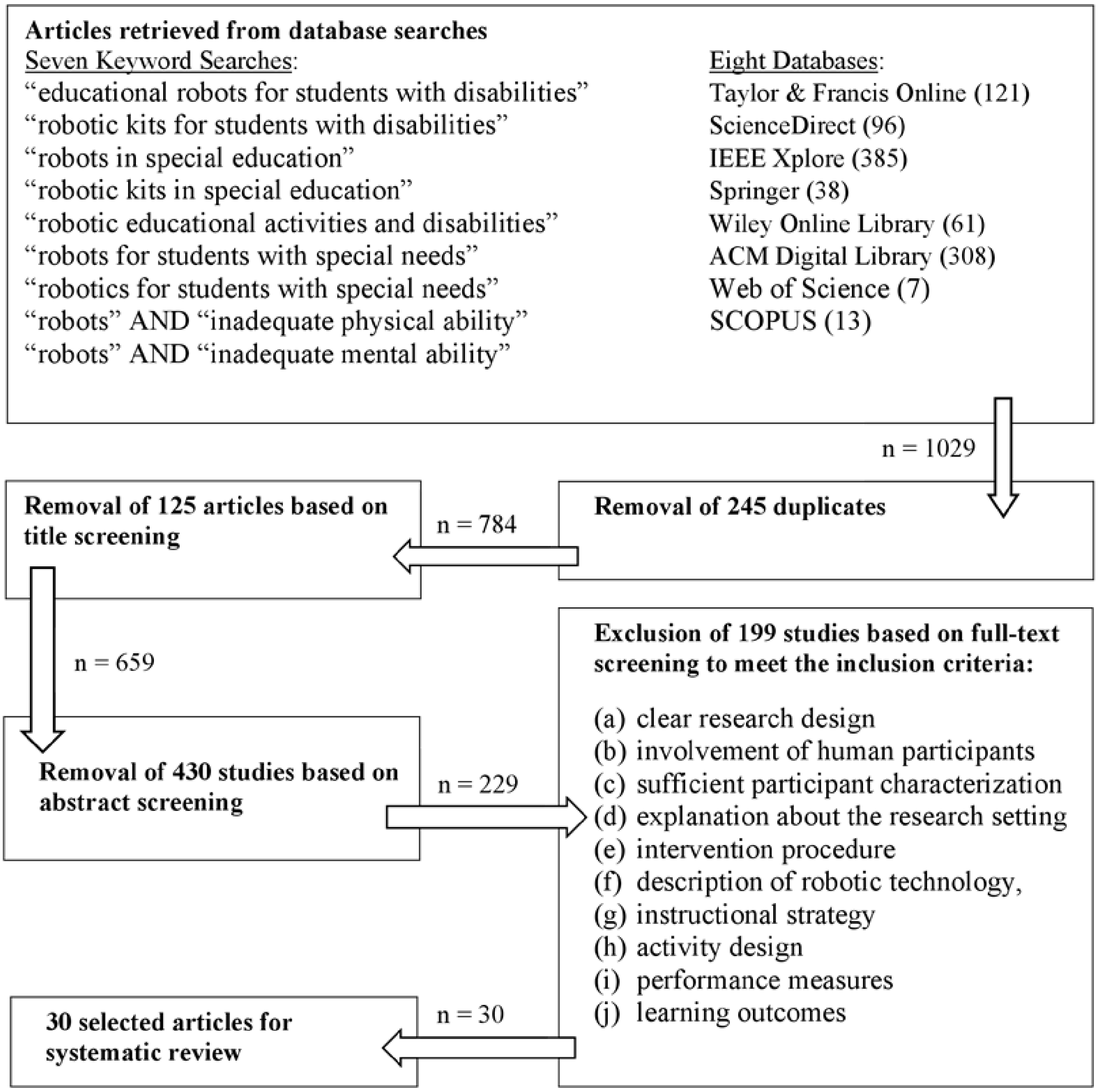


Figure 1. Flow chart for the article search and selection process

2.2. Research rigor and design of the selected studies

The researchers assessed the research rigor using *Horner's Criteria,* a widely adopted rubric for case design in special education (Moeller et al., 2015). This set of criteria was valid because case design is commonly used in special education research, and a large proportion of the selected studies employed case design. To ensure the credibility of Horner's quality indicators, the researchers checked them against the essential quality indicators for experimental research in special education (Gersten et al., 2005) and found that Horner's Criteria sufficiently fulfilled items for Describing Participants, Implementation of the Intervention and Description of Comparison Conditions, and Outcome Measures. Horner's criteria further assessed Social Validity of case design (Moeller et al., 2015). In comparison to ideal indicators for measuring qualitative methodological rigor in general education research, Horner's criteria adequately ensured Responsiveness to Social Context, Appropriateness of Sampling, Adequacy of Sampling, and Transparency of Data Collection (Fossey et al., 2002). As shown in Table 1, low percentages were reported for establishing baseline conditions (33.3%) and ensuring experimental control (46.7%), implying challenges faced by special education practitioners when designing and implementing robotic interventions. In terms of the research design, single-group interventions (e.g., summer camps) measured by pre- and post-tests (n = 14) and case studies (n = 9) were adopted most frequently (See Table S1 in Supplementary Materials).

Table 1. Rigor assessment of the robot-assisted special education studies

Quality indicators	Number of studies meeting the criteria	% of studies meeting the criteria
1. Participants & Settings		
Participant description	30	100%
Participant selection/recruitment	26	86.7%
Setting description	30	100%
2. Dependent Variable (Outcome)		
Operationally defined	30	100%
Measurement of performance is valid and described sufficiently	29	96.7%
Inter-observer agreement or strict confirmability checks	15	50%
3. Independent Variable (Treatment)		
Operationally defined	29	96.7%
Systematically manipulated by experimenter	30	100%
Implementation fidelity established	30	100%
4. Baseline		
Baseline conditions are operationally defined	10	33.3%
5. Internal Validity		
Controlled for common threats to internal validity	15	50%
Demonstrated experimental control	14	46.7%
6. External Validity		
Experimental effects are replicated across participants, settings, or materials	23	76.7%
7. Social Validity		
Dependent variable is socially important	29	96.7%
Magnitude of change is socially important	29	96.7%
Implementation is practical and effective	29	96.7%

2.3. Using Activity Theory to analyze robot-assisted special education studies

This review adopted Activity Theory to perform content analysis on the interplay of various components and actors in robot-assisted special education research. Since activity was defined as a system of purposeful behaviors leading to recognizable changes in human practices (Kim, 2010), the researchers examined how robot-assisted learning could lead to evolving behaviors and practices among the stakeholders. As shown in Figure 2, the framework addressed how robot products were adapted for learning, how the robotic mediation led to learning outcomes, how special education professionals created and perceived the learning environment enriched by robotic activities, and how parents were involvement in the robot-assisted interventions.

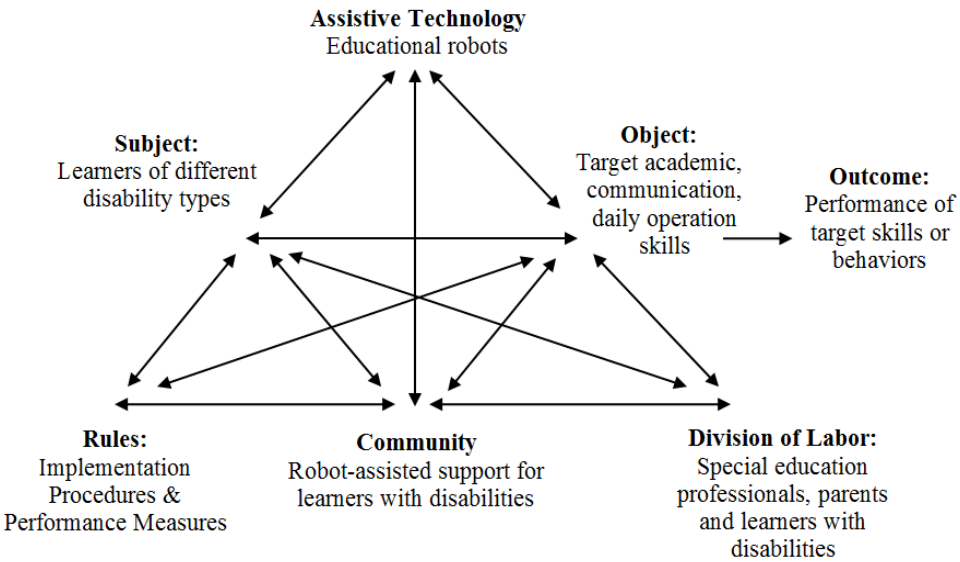


Figure 2. Using Activity Theory to analyze robot-assisted special education studies

Subject referred to learners with one or more different disabilities who participate in robot-assisted special education research; *Technology* referred to educational robots, the accompanying tools (e.g., haptic and sound generation devices) and instructional design strategies that mediated the learning. *Object* included skills and behaviors that the robotic technology aimed to improve among learners (e.g., academic, communicative, social/interactional, movement, daily operation, executive functions and perceptions, active task engagement). *Rules* included accepted practices in implementing robot-assisted interventions, for example, steps in the intervention procedures and performance measures for evaluating learning outcomes; and *Community* referred to people involved in robot-assisted interventions (e.g., learners, family, friends, professionals) and special education settings (e.g., schools, clinics) that supported the interventions. *Division of Labor* referred to distribution of duties among learners, special education professionals, and parents for undertaking robot-assisted learning interventions. *Outcome* was learners' performance in target skills as evaluated by performance measures. Two of the researchers iteratively analyzed the content from Tables S1 and S2 and identified instances based on this coding scheme.

2.4. Categorization of learner disabilities

During the content analysis, two of the researchers first summarized the participant profile, disability categories, learning domain and objectives, robot type, accompanying tools, research design, instruments, learning activity, role of robot, instructional strategy, performance measures, and outcomes (See Supplementary Materials for Tables S1, S2, and S3). The disabilities in each study were then grouped into six categories based on disability dimensions and types listed by the International Classification of Functioning and Disability, and Health (Perenboom & Chorus, 2003) and the Individuals with Disabilities Education Act (Office of Special Education and Rehabilitation Services, 2020). The criteria for the categorization included disability dimensions (e.g., mobility, social integration) and disability types (e.g., Intellectual Disability, Hearing Impairment) that repeatedly surfaced in the coding. When there was a discrepancy, the two coders discussed based on the established criteria and reached consensus on the final coding results.

The first category included Emotional or Attention-Related Disabilities (e.g., ASD, ADHD) involving diagnosed problems with neurodevelopment, behaviors, and communication. The second was Intellectual Disability, which referred to a general learning disorder that affected their intellectual and adaptive functioning. The third was Physical Disabilities that affected movement and physical development, and the fourth was Sensory Impairments of eyesight or hearing. The fifth was Speech Impairments, which affected one's ability to communicate, and the sixth was Unspecified Disabilities that caused learning difficulties. The coders also grouped performance measures among the studies into categories. Table S2 in Supplementary Materials shows the seven performance measure categories ranked by frequency based on the coding scheme in Table 2. It should be noted that several studies used more than one type of performance measure. Finally, the role of robot and the learning objectives were analyzed based on information provided in Tables S1 and S2 in Supplementary Materials.

Table 2. Coding Scheme for Performance Measures

Code	Description
Academic Performance	• Learning outcomes in academic subjects (e.g., STEM, art)
Communication	• Expressive and receptive use of language (e.g., oral and listening skills) • Ability to use augmentative and alternative communication technology
Active Task Engagement	• Taking initiatives to participate and staying focused during task • Awareness, attention, curiosity, persistence throughout task
Social/Interactive Skills	• Ability to carry out activities with others (e.g., robot, peers, or adults)
Executive/Perceptual Functions	• Executive functions (e.g., working memory) • Visual, auditory, haptic, spatial skills
Kinesthetic Behaviors	• Physical movements and motor skills • Daily operations and functions
Interest/Motivation/Attitude	• Interest in using assistive technology for robotic programming • Desire to participate in learning activities • Willingness to complete learning activities

3. Results and discussion

To address the first research question, findings based on the 30 studies showed the design features of learning activities for different disability categories supported by different hardware and instructional strategies, research implementation processes enacted by various stakeholders, and outcome evaluation in robot-assisted special education. To address the second research question, recommendations on how to improve robot-assisted special education were provided based on the identified gaps, challenges, or contradictions under each activity component.

3.1. Subject

The subjects were learners between 2 and 21 years of age with one or more disabilities. Figure 3 shows that fourteen studies focused on learners with Emotional or Attention-Related Disabilities. These learners mostly suffered from ASD and needed support on social interaction and communication (e.g., Albo-Canals, 2018). Ten studies focused on Intellectual Disability, particularly those with moderate or mild intelligent quotients, and inadequate general thinking abilities (e.g., cause-effect conceptualization) and poor sense of direction (Bargagna et al., 2018; Lee & Hyun, 2015; Park & Kwon, 2016; Pennington et al., 2014). Nine studies involved learners with Physical Disabilities. Four and three studies concerned learners with Sensory Impairments and Speech Impairments respectively. Learners with Sensory Impairments had little experience with programming (e.g., Howard et al., 2012), while learners with Speech Impairments had below-average expressive and receptive language levels (Encarnação et al., 2017; Lee & Hyun, 2015). Finally, two studies involved learners with Unspecified Disabilities displayed deficiency in reading, writing, and self-directed functions (Karna-Lin et al., 2006; Pihlainen et al., 2017).

In each study, learners received specific support and training according to the disability categories reported in Figure 3. For instance, learners with Physical Disabilities received support on overcoming severe physical limitations in order to engage in robotic play (Adam & Cook, 2013; Cook et al., 2005; Van den Heuvel et al., 2017a; Van den Heuvel et al., 2017b; Van den Heuvel et al., 2020). The main challenge for the Subject component was heterogeneity in the learners' disability profiles. Due to this great variety in disability types, each study designed its own robot-assisted learning activities and assessment methods to meet the learners' specific conditions and needs.

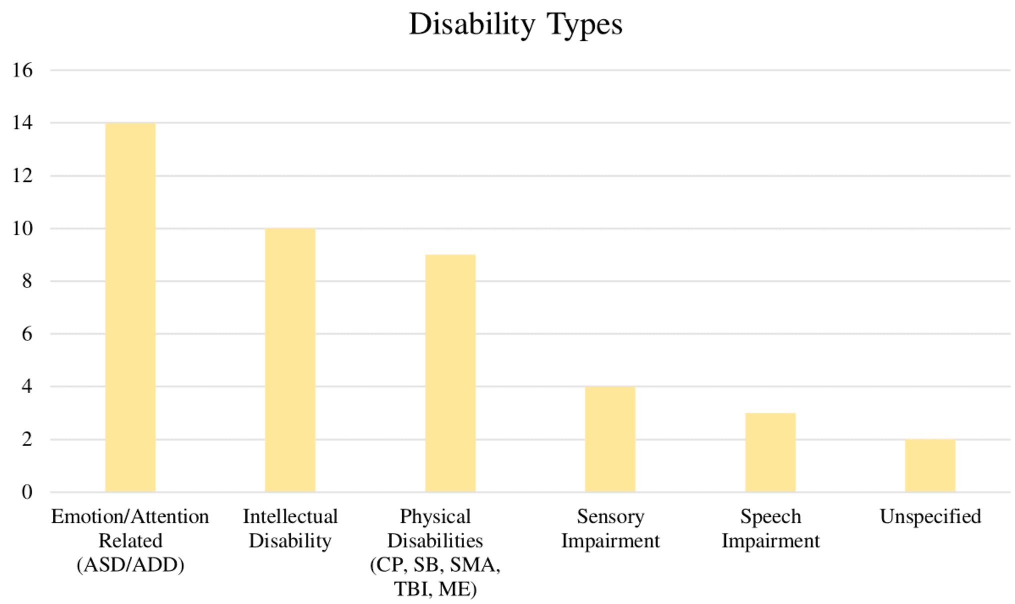


Figure 3. Categorization of learner disabilities in the 30 reviewed studies

Recommendations: To improve the activity system for the Subject component, a repository can be created to offer practical guidelines on how to design robot-assisted learning activities for each disability category shown in Figure 3. Such categorization will facilitate the role of special education practitioners in designing the appropriate robotic learning activities for each disability category. For example, the activity design strategies and tools for learners with Physical Disabilities could differ greatly from those for ASD learners. Clustering the design strategies thus ensures that practitioners can have a simple and efficient design and implementation experience using robots as mediating tools to meet the needs of learners in each of the six disability categories we identified.

3.2. Technology

The *Technology* component was divided into hardware and instructional design. The hardware consisted of three forms of robots — (a) commercial robot kits (e.g., LEGO Mindstorms), (b) humanoids (e.g., NAO NexGen), and (c) manipulatives (e.g., Bee Bot). In terms of how the hardware was utilized in the designed activities, robot kits were used as an animal or vehicle character whose movements or behaviors were programmed and designed by the participants, whereas humanoids served as either a learning companion or an in-class tutor.

Figure 4 shows that the most common use of robots (50%) was as a vehicle, animal, or manipulative designed and programmed by learners as a technological product (e.g., Wright, Knight, Barton, & Edwards-Bowyer, 2019). The robots were manipulated through different channels such as sound generating device or switches (e.g., Bargagna et al., 2018; Disseler & Mirand, 2017). The second use of robots was a learning companion (47%), where humanoids acted as playable peers who trained learners' academic skills (e.g., Cook et al., 2005; Freitas et al., 2017), communication (e.g., Lee & Hyun, 2015; Saadatzi et al., 2018), and social and interactional capabilities (e.g., Albo-Canals et al., 2018; Huijnen, Lexis, Jansens, de Witte, 2016). The humanoids also achieved therapeutic effects on movements (e.g., Khaksar et al., 2019; Van den Heuvel et al., 2017b) and daily functions (e.g., Park & Kwon, 2016). The third use of robots was a humanoid in-class tutor (3%) that instructed learners how to communicate appropriately through text messaging (Pennington et al., 2014).

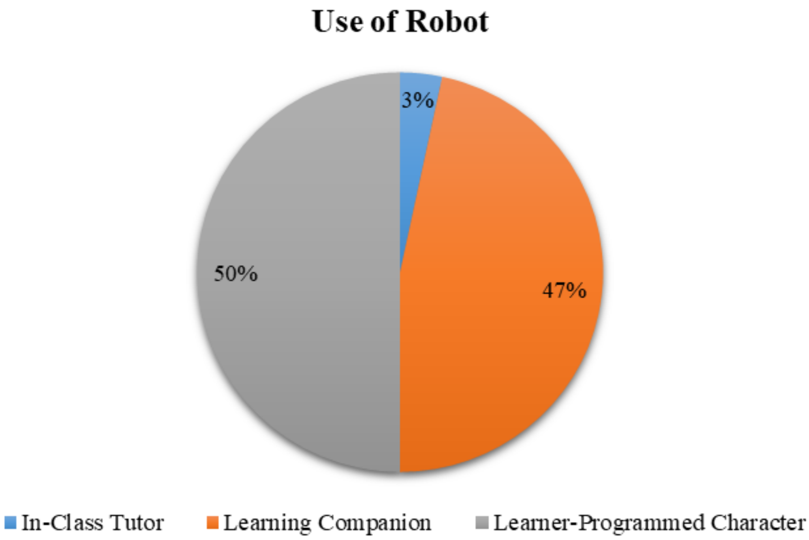


Figure 4. The role of robot in the 30 reviewed studies

The use of robots was consolidated by domain-specific instructional strategies and activities. First, the instructional design of robotic programming employed strategies, such as problem solving (e.g., Karna-Lin et al., 2006; Ludi & Reichlmayr, 2011; Yuen, Mason, & Gomez, 2014), multimodal interaction and feedback (e.g., Howard et al., 2012, Dorsey et al., 2013), competition (e.g., Dorsey et al., 2013; Howard & Park, 2014), and inquiry (e.g., Disseler & Mirand, 2017; Jung, Lee, Cherniak, & Cho, 2019) to enhance the learning process and heighten learners' motivation. Personalized adaptations strategies were also used to accommodate learners' needs (Lindsay & Hounsell, 2016). For instance, physical adaptations made use of a magnifying app for visually impaired learners; cognitive adaptations included the use of prompts, reminders, and simplified instruction in reading; and social adaptations included anxiety reduction and pairing of cooperative learners (Lindsay & Hounsell, 2016). Lastly, inquiry (Disseler & Mirand, 2017) and interactional games were used for academic learning (Freitas et al., 2017).

Strategies for language and literacy development included scripted play and talk (Lee & Hyun, 2015; Van den Heuvel et al., 2017a; Van den Heuvel et al., 2017b); affective learning (Lee & Hyun, 2015); self-learning (Lee & Hyun, 2015); robot therapy (Lee & Hyun, 2015); fun elicitations (Huijinen et al., 2016); and social interactional learning made use of picture-verb matching and picture-sentence matching (Park & Kwon, 2016). To train executive functions and visuo-spatial skills, strategies such as learning-by-doing and learning-by-thinking were used to enhance visual-working memory and intellectual abilities (Bargagna et al., 2018). Finally, to improve general functions and motor skills, social interactional learning and play-like activities were employed to improve movement and social interaction skills (Park & Kwon, 2016; Van den Heuvel et al., 2017b).

Table 3 shows the five derived learning domains and activity types in robot-assisted special education. Robotic Design and Programming activities (e.g., Disseler & Mirand, 2017; Howard, Park, & Remy, 2012; Ludi & Reichlmayr, 2011; Pihlainen et al., 2017) included building a robot (Karna-Lin et al., 2006) programming for robotic movements (Bargagna et al., 2018) and executions (Howard et al., 2012), connecting modules (Adams & Cook, 2013), teaching robots to draw (Howard & Park, 2014), and constructing robot cars (Disseler & Mirand, 2017). Activities focusing on the design of robotic actions provided participants the experience technological development through co-ideation with adults (Bertel et al., 2013) or designing challenges (e.g., moving down a maze; Ludi & Reichlmayr, 2011). The design and programming activities also effectively increased the participants' motivation, concentration, communication, and activeness as they learned to create interesting technology constructs (Karna-Lin et al., 2006).

Robotic Play activities made use of different robotic roles to create interaction effects. A child-size and doll-like humanoid helped autistic children participate in multi-modal social interactions (Huijnen et al., 2016). A companion robot also successfully provided learners with interactive scenarios for turn-taking, goal achievement, and sensory interactions (Van den Heuvel et al., 2017a). Similarly, a humanoid provided therapy and communication skill building through robotic-play scenarios that enabled robotic control through vocal commands or sensor pressing (Van den Heuvel et al., 2017b) or through telepresence in classroom or home settings for fostering relatedness (Culen et al., 2019).

For Literacy Development in science, math, languages, and social studies, learners practiced sight words with a truck-like robot character (Saadatzi et al., 2018) through Question and Answer, Counting, and Sequencing Events during story reading (Encarnação et al., 2017; Jordan et al., 2013). Exploration and Ideation activities, on the other hand, made use of the Bee Bot Robot and Cubelets (Jung, Lee, Cherniak, & Cho, 2019) to help high-functioning ASD learners participate in non-sequential inquiry. The sensor-embedded Cubelets are manipulative blocks assembled and programmed by learners with ASD. Additionally, a therapeutic humanoid was used for co-ideation and co-creation in parent-child participatory learning (Bertel et al., 2013).

Kinesthetic Tasks focused on motions, gestures, and dance. Humanoids (e.g., ZORA and Nao NextGen) were used to facilitate the kinesthetic development via leg movement exercises and dancing with songs (Van den Heuvel et al., 2017b). The robots also rewarded learner behaviors, provided learning cues, and induced active learning. The kinesthetic activities engaged learners more effectively than conventional classroom activities (Hedgecock et al., 2014). Finally, Artistic Creation activities used LEGO robotic products. Learners created music by drawing on a screen using the LEGO Education EV3 Mindstorms Kit (Pihlainen et al., 2017).

Table 3. Learning domains and activity types supported by different robotic hardware

Learning domains	Activity type					
	Robotic Design & Programming (n = 20)	Robotic Play (n = 14)	Literacy Development (n = 8)	Exploration & Ideation (n = 4)	Kinesthetic Tasks (n = 3)	Artistic Creation (n = 2)
Academic Development in STEM or Humanities (n = 17)	n = 10 Robot Kits	n = 3 Humanoids	n = 1 Robot Kits	n = 3 Humanoids Manipulatives		
Communication & Social/Interactional Skills (n = 17)	n = 3 Robot Kits	n = 7 Humanoids	n = 5 Humanoids		n = 2 Humanoids	
General Operation & Movement (n = 9)	n = 2 Robot Kits	n = 3 Humanoids	n = 2 Humanoids	n = 1 Humanoids		n = 1 Manipulatives
Executive Functions & Perceptual Skills (n = 6)	n = 2 Robot Kits	n = 2 Humanoids			n = 1 Humanoids	n = 1 Manipulatives
Active Task Engagement (n = 2)	n = 3 Robot Kits					

Recommendations: Robot-assisted interventions should implement more Exploration and Ideation, Kinesthetic, and Artistic Creation activities (See Table 3) because these types of activities can help learners in special education develop cognitively, emotionally, and behaviorally (Denisova et al., 2019). A recent study on teacher training further shows the importance of implementing creative activity design integrating robots in special education (Coskun, 2020). To fill this gap, more efforts should be devoted to investigate ways to foster learners with disabilities through creative tasks and kinesthetic engagement. For instance, more teacher training should be provided to improve the design and use of robotic learning activities in special education for better learning outcomes. Moreover, guidelines that aim to improve domain-specific skills (e.g., Academic Development or General Operation and Movement) should be offered for each of our proposed disability category in a repository of standard practices for robot-assisted special education.

3.3. Object

The objectives of the reviewed studies were to train learners' (a) academic skills in STEM, (b) communication and social interaction skills, (c) general operation and movement, (d) executive functions and perceptual skills, and (e) active task engagement. Table 3 shows that while many studies focused on training academic, communication, and social interaction skills, few studies investigated learners' active task engagement (n=3). This finding points to the need to shift the focus of research objectives in robot-assisted special education toward self-initiated learning.

Recommendations: Robot-assisted special education practitioners should formulate more learning objectives to increase active task engagement and create a positive impact on the career readiness for learners with disabilities. This will help them gain self-efficacy and positivity when going into a career related to computers and robotic technology (Ludi & Reichlmayr, 2011). One way to achieve this is to incorporate group-based, hands-on activities during workshops and camps with designs that allow them to share ideas and learn from one another.

3.4. Rules

The *Rules* component focused on Intervention Procedure and Performance Measures. The studies consistently followed strict implementation procedures, including (a) recruitment of target learners based on selection criteria, (b) training the professionals on using robotic technology for instruction, (c) training the learners how to use the robots in the learning activities, (d) establishing baseline conditions, (e) designing the experiment or case study, (f) implementing the learning activities, and (g) evaluating the learners' performance based on well-defined measures.

It was essential to apply precise quantitative and qualitative Performance Measures in robot-assisted special education due to challenges, such as the heterogeneous nature of participant profiles, small sample sizes, and the absence of control groups in many of the studies. Both objective (e.g., Howard, Park, & Remy, 2012) and subjective (e.g., Dorsey et al., 2013; Howard et al., 2012; Karna-Lin et al., 2006) measures were used to evaluate learners' goal attainment (Cook et al., 2005), engagement (Hedgecock et al., 2014), language development (Encarnação et al., 2017), interest and efficacy (Ludi & Reichlmayr, 2011), or neuropsychological and cognitive development (Bargagna et al., 2018).

Based on the coding scheme in Table 2, seven types of Performance Measures were identified (See Figure 5), Academic Performance ranked first and measured STEM performance in robotic design and programming (e.g., Howard et al., 2012; Jung et al., 2019), mathematics (Encarnação et al., 2017; Freitas et al., 2017), and physics (Disseler & Mirand, 2017). Active Task Engagement ranked second. This measure concerned the extent to which learners actively participated in the activities and stayed focused on task with persistence (e.g., Freitas et al., 2017; Hedgecock et al., 2014; Karna-Lin et al., 2006). Communication/Language Use and Social/Interactive Performance ranked equally as the third most commonly adopted measures. Communication/Language Use focused on skill performance in conversation, word use, and Question-and-Answer (e.g., Encarnação et al., 2017; Lee & Hyun, 2015); and Social/Interactive Performance assessed participation during play, turn-taking, and collaboration in social interactions (e.g., Albo-Canals et al., 2018; Huijnen et al., 2016; Van den Heuvel et al., 2017a).

The Rules component bears several challenges. First, since every learner had different disabilities, it was difficult to maintain a single set of Intervention Procedures or Performance Measures. Second, many learners could not directly communicate their user experience and feedback, the practitioners could not collect their real experience and opinions about the robotic interventions. Third, in terms of outcome evaluation, traditional pen-and-paper tests have been used to assess the learners in the reviewed robot-assisted special education studies. However, the lengthy process of these tests may discourage learners with disabilities (Disseler & Mirand, 2017).

Recommendations: Special education practitioners should explore the potential of robots as evaluators or assessment tools, and use them in creating new implicit methods for assessing outcomes in special education assisted by educational robots. For instance, it is possible to make robots act as evaluator by collecting log history of learners' interactions and analyze them to provide assessment on active task engagement, motivation, and learning outcomes. Based on the assessment results, it is possible to offer adaptive learning paths in robot-assisted special education to learners in a specific disability category. This is a potential research niche, as no previous study has reported the use of adaptive learning in robot-assisted special education.

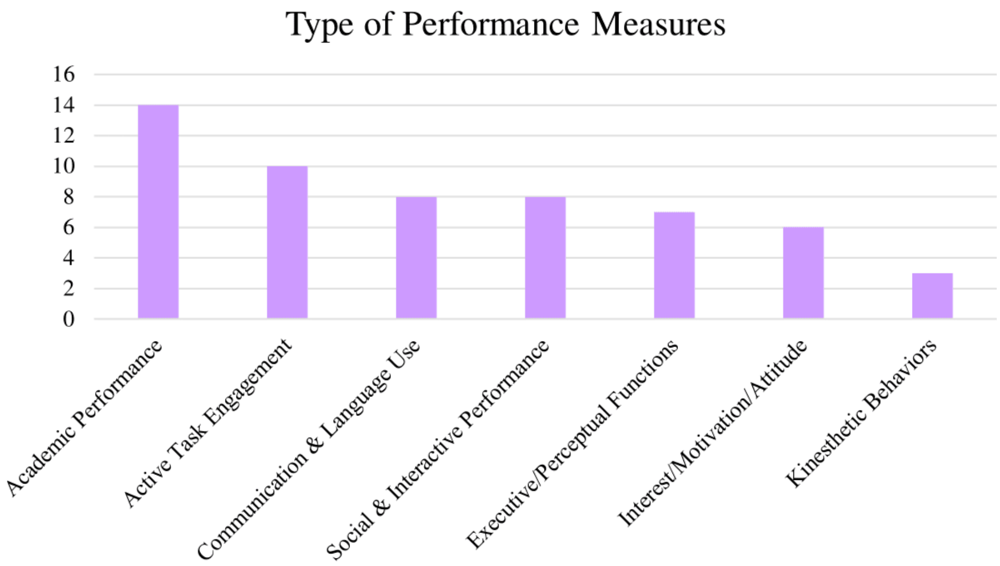


Figure 5. Classification of performance measures used in the 30 reviewed studies

3.5. Community

The *Community* component included special education professionals, parents, family, friends/classmates, and learners with disabilities. All of the reviewed studies involved special education professionals, namely, therapists, researchers, educators, facilitators, or designer in the interventions, however, only three studies involved the participation of parents or family members in the interventions (Pihlainen et al., 2017; Lindsay & Hounsell, 2016); and only one study involved friends or classmates of adolescents with Chronic Fatigue Syndrome (Culen et al., 2019). This shows a need for more parental participation in robot-assisted special education as the participation of parents in these activities can lead to better learning experience and outcomes (e.g., Lindsay & Hounsell, 2016). As for the variety of special education settings, the interventions were carried out at special education general schools or special education schools (*n* = 16), rehabilitation/care centers or hospitals (*n* = 9), robotics camps/workshops (*n* = 3), educational technology laboratories (*n* = 2), technology clubs (*n* = 1), or at home (*n* = 1).

A number of challenges for the Community component were identified. First, some robot-assisted intervention settings did not adequately fulfill professional and parental expectations of a safe, pleasant, and inclusive learning environment. Second, when interventions were conducted in an inclusive environment, real constraints were reported by mainstream teachers who practiced inclusion education (Encarnação et al., 2017), including difficulty with managing the different time frames between typically developing and learners with disabilities during robotic learning activities. Third, the adaptability of the mediating environment was considered low, and higher robot autonomy was expected in therapeutic scenarios.

Recommendations: In response to the challenges, several recommendations have been put forth. First, designers of robot-assisted interventions should involve parents during the process of creating a safe and pleasant environment for learners with disabilities to work toward their professional goals. This would help the participants live more independently in the future (Huijnen et al., 2016). Second, with the goal to foster an inclusive environment, robot-mediated learning environments should provide not only educational benefits, but also social inclusion. This would ensure long-term success of the learners with disabilities in the mainstream society (Ludi & Reichlmayr, 2011). Third, robots should have adaptive functionalities to meet individual learner needs along the intervention timeline. It would be desirable to design robots based on users' attitudes, perceived adaptability, perceived usefulness, intention to use, perceived enjoyment, and trust (Huijnen et al., 2016). Finally, concerning organizational capacity, more research should be conducted to investigate the extent to which learning institutions and schools can accommodate the needs of learners with disabilities.

3.6. Division of labor

The *Division of Labor* was among (a) learners with disabilities, (b) special education professionals, and (c) parents. Firstly, the learners were involved in different ways by (a) receiving training on using robots to complete learning activities, (b) participate in robot-assisted learning activities to improve specific skills or knowledge, and (c) providing individual feedback on robot-assisted learning. Second, special education professionals engaged in (a) recruiting participants, (b) training the professionals and the learners, (c) designing and implementing the intervention/learning activities, (d) facilitating learners throughout the robot-assisted learning process, (e) creating performance measures and evaluating the learning outcomes, and (f) providing feedback on the use of robots in special education. Third, parents were involved in different ways. Only in a few studies did they play active roles in the design of robotic solutions by providing support on technology development and feedback on the effectiveness of the robot-mediated instruction (e.g., Pihlainen et al., 2017). In some studies, parents participated by looking at the learners' designed products (Lindsay & Hounsell, 2016) or taking care of them prior to, during, and after the robot-mediated learning process.

A gap was identified with respect to learners' participation in the design of learning activities and interventions. No study that involved the learners with disabilities in the design or planning process for the robot-assisted interventions. This was a missing element that might have affected the suitability of the designed activities and implementation procedure, as it was necessary to cater to the learners' disability differences in special education (Encarnação et al., 2017; Huijnen et al., 2016).

Recommendations: In terms of instructional design and implementation, more responsibilities should be shifted to parents and learners in the activity system. The parents can be more actively involved in the intervention process and play active roles such as co-designing a programmable character using robot kits. Special education practitioners who plan to design robot-assisted interventions can also invite learners with their target disability types to participate in pilot tests to help practitioners improve their activities before actual interventions. Such co-creation and co-participatory design of target robotic learning activities will make the learning experience beneficial for the learners.

3.7. Outcome

Outcomes based on specific performance measures are discussed based on disability categories and perceptions of special education professionals. Specifically, educational robots made learners with Emotion/Attention Related disabilities (ASD, ADD) more motivated and engaged in learning. They also helped them acquire knowledge from several domains, including cause-effect concepts, robotic programming skills, and acquisition of sight words. Additionally, educational robots enhanced the communication and social interaction of learners with Emotion/Attention Related disabilities. For learners with Intellectual Disability, not only did educational robots motivate them, but also they enhanced their knowledge, skills and intelligence quotient.

For learners with Physical Disabilities, educational robots helped them become more physically active through body movement tasks. Robots also facilitated these learners' goal attainment, as well as their movement, communication and robot programming skills. Learners with Speech Impairments, on another note, perceived learner-robot interactions as effective language learning treatment. Moreover, educational robots made learners with Sensory Impairments highly interested and confident in learning. For instance, visually impaired learners improved their programming as well as visual skills through maze navigation (Dorsey et al., 2013). Finally, for learners who suffered learning difficulties due to Unspecified Disabilities, educational robots enhanced their learning motivation, concentration, social, and teamwork skills.

Perceptions of special education professionals revealed the effectiveness and benefits of adapting the robotic program for the needs of target learners (e.g., Lindsay & Hounsell, 2016). The professionals approved of the suitability of the robotic play content related to daily life and the valuable potential of using specific educational robots (e.g., KASPAR) for working on therapeutic and educational goal attainment (Huijnen, Lexis, & de Witte, 2016).

One contradiction concerned whether robots should replace humans. Many special education experts expressed their ethical consideration about the use of robots as a social interactive agent. They disapproved of the absence of a human therapist in robot-mediated environments (Lee & Hyun, 2015), indicating the inability of robots to provide natural responses outside of script content. Additionally, different views on whether the robot-mediated environment adequately fulfilled their expectations of a safe, pleasant, and inclusive learning environment were identified (Huijnen et al., 2016). Another issue was the amount of time and effort spent on choosing suitable robots based on their appearance in order to appeal to learners with disabilities.

Recommendations: First, researchers should investigate appropriate appearances of educational robots (e.g., human-like, toy-like, or animal-like) and variability in their facial expressions based on the needs of different disability categories. Adapting the robot to individual learner needs along the intervention timeline is also recommended. Special education practitioners should create robot-assisted special education standards about the treatment frequency, duration, and robots' roles in scaffolding social interactions to ensure that robotic interventions scaffold social interaction skills among learners (e.g., ASD learners) toward interactions with real humans. Moreover, it would be desirable to design the robots based on users' attitudes, perceived adaptability, perceived usefulness, intention to use, perceived enjoyment, and trust (Encarnação et al., 2017; Huijnen et al., 2016). Professional training for instructors and staff on such matters will help to improve robot-assisted special education (Huijnen et al., 2016).

4. Conclusion

In order to understand the process involved in robot-assisted special education and the role played by its various stakeholders, this systematic review probed into the design, implementation, and outcome of robot-assisted special education research through the perspective of Activity Theory. Major components of the robot-assisted learning activity system were analyzed, including (a) learners with disabilities, (b) robots supported by instructional design, (c) target skills, (d) intervention procedure and performance measures, (e) community of special education professionals, parents, and learners, (f), division of labor among learners, special education professions, and parents, and (g) learning outcomes. The analysis showed that practitioners need to align robot-mediated instructional design with disability categories of target learners, learning strategies, and robot types. This connection among the activity components would create more effective learning and generate greater benefits for learners with disabilities. The review further provided recommendations based on each activity component so that existing challenges, gaps, and contradictions can be minimized in future design and implementation of robot-assisted interventions in special education.

Several limitations should be acknowledged. First, the review consisted of a small number of studies due to the stringent inclusion criteria. Although many articles were retrieved initially, after full-text screening, only 30 studies qualified as our data. However, the small set of data allowed us to reach saturation during the coding to sufficiently address the research questions. Another limitation was the absence of gender as a factor. Only a few studies (e.g., Encarnação et al., 2017; Van den Heuvel et al., 2017b) provided information about the gender of the learners.

The findings contribute to the research fields of special education and robotic technology by providing directions for research design and implementation of robotic-assisted special education. Specifically, more robot-assisted activities should aim to foster creative and kinesthetic skills. Future research may aim to (a) develop general guidelines across learning domains in addition to domain-specific guidelines (e.g., skills in STEM or Humanities, Social/Interactional Skills) for each disability category (e.g., Emotion/Attention Related, Intellectual Disabilities, Physical Disabilities, Social/Interaction Disabilities) and b) designing career preparation activities to help learners with disabilities build an autonomous future life.

Acknowledgement

This research was supported by the Ministry of Science and Technology, Taiwan under project numbers MOST-108-2511-H-224-008-MY3 and MOST-107-2511-H-224-007-MY3.

Reference

108th Congress. (2004). *Public Law 108–446: Individuals with Disabilities Education Act of 2004*. Retrieved from https://ies.ed.gov/ncser/pdf/pl108-446.pdf

*Adams, K., & Cook, A. (2013). Programming and controlling robots using scanning on a speech generating communication device: A case study. *Technology and Disability, 25,* 275–286. doi:10.3233/TAD-140397

*Albo-Canals, J., Martelo, A. B., Relkin, E., Hannon, D., Heerink, M., Heinemann, M., Leidl, K., & Bers, M. U. (2018). A Pilot study of the KIBO Robot in children with severe ASD. *International Journal of Social Robotics*, *10*(3), 371-383. doi:10.1007/s12369-018-0479-2

Alcorn, A. M., Ainger, E., Charisi, V., Mantinioti, S., Petrović, S., Schadenberg, B. R., Tavassoli, T., & Pellicano, E. (2019). Educators' views on using humanoid robots with autistic learners in special education settings in England. *Frontiers in Robotics and AI,* 6, 1-15. doi:10.3389/frobt.2019.00107

Ashton, T. M. (2005). Assistive technology [Editorial]. *Journal of Special Education Technology, 20*(2), 60-63. doi:10.1177/016264340502000208

*Bargagna, S., Castro, E., Cecchi, F., Cioni, G., Dario, P., Dell'Omo, M., Di Lieto, M. C., Inguaggiato, E., Martinelli, A., Pecini, C., & Sgandurra, G. (2018). Educational robotics in down syndrome: A Feasibility study. *Technology, Knowledge and Learning, 24*(2), 315-323. doi:10.1007/s10758-018-9366-z

Benitti, F. B. V. (2012). Exploring the educational potential of robotics in schools: A Systematic review. *Computers & Education, 58*(3), 978-988. doi:10.1016/j.compedu.2011.10.006

*Bertel, L. B., Rasmussen, D. M., & Christiansen, E. (2013). Robots for real: Developing a participatory design framework for implementing educational robots in real-world learning environments. In P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, & M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2013, Lecture Notes in Computer Science, 8118* (pp. 437-444). Berlin, Germany: Springer. doi:10.1007/978-3-642-40480-1_29

Catlin, D., & Blamires, M. (2018). Designing robots for special needs education. *Technology, Knowledge and Learning*, *24*(2), 291-313. doi:10.1007/s10758-018-9378-8

Chang, C.-W., Lee, J.-H., Chao, P. Y., C., Wang, C. Y., & Chen, G. D. (2010). Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. *Educational Technology & Society*, *13*(2), 13-24.

Cheng, Y. W., Sun, P. C., & Chen, N. S. (2018). The essential applications of educational robot: Requirement analysis from the perspectives of experts, researchers and instructors. *Computers & Education, 126*, 399-416. doi:10.1016/j.compedu.2018.07.020

*Cook, A. M., Bentz, B., Harbottle, N., Lynch, C., & Miller, B. (2005). School-based use of a robotic arm system by children with disabilities. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 13*(4), 452-460. doi:10.1109/TNSRE.2005.856075

*Culen, A. L., Borsting, J., & Odom, W. (2019). Mediating relatedness for adolescents with ME: Reducing isolation through minimal interactions with a robot avatar. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (pp. 359–371). doi:10.1145/3322276.3322319

Daniels, H., & Cole, T. (2002). The Development of provision for young people with emotional and behavioural difficulties: An Activity theory analysis. *Oxford Review of Education, 28*(2 & 3), 311–329.

Datteri, E., Zecca, L., Laudisa, F., & Castiglioni, M. (2013). Learning to explain: The Role of educational robots in science education. *Themes in Science & Technology Education*, *6*(1), 29-38.

Denisova, O. A., Gudina, T. V., Bukina, I. A., Lekhanova, O. L., & Selina, A. V. (2019). Formation of adaptive and integrative potential of children with mental retardation in artistic and creative activities. *The Education and Science Journal, 21*(8), 164-186. doi:10.17853/1994-5639-2019-8-164-186

*Disseler, S., & Mirand, G. (2017). Students with disabilities and LEGO© education. *Journal of Education and Human Development*, *6*(3), 38-52. Retrieved from http://jehdnet.com/journals/jehd/Vol_6_No_3_September_2017/5.pdf

*Dorsey, R., Park, C. H., & Howard, A. (2013). Developing the capabilities of blind and visually impaired youth to build and program robots. *Journal on Technology and Persons with Disabilities, 1*, 57-69. Retrieved from http://scholarworks.csun.edu/bitstream/handle/10211.3/121965/JTPD201406-p57-69.pdf?sequence=1

Edwards, A., Gilroy, P., & Hartley, D. (2002). *Rethinking teacher education: Collaborative responses to uncertainty*. London, UK: RoutledgeFalmer.

Engeström, Y. (2001). Expansive learning at work: Toward an activity-theoretical reconceptualization. *Journal of Education and Work*, *14*(1), 133–156. doi:10.1080/13639080020028747

*Encarnação, P., Leite, T., Nunes, C., Nunes da Ponte, M., Adams, K., Cook, A., Caiado, A., Pereira, J., Piedade, G., & Ribeiro, M. (2017). Using assistive robots to promote inclusive education. *Disability and Rehabilitation: Assistive Technology, 12*(4), 352-372. doi:10.3109/17483107.2016.1167970

Fossey, E., Harvey, C., McDermott, F., & Davidson, L. (2002). Understanding and evaluating qualitative research. *Australian & New Zealand Journal of Psychiatry, 36*(6), 717-732. doi:10.1046/j.1440-1614.2002.01100.x

*Freitas, H., Costa, P., Silva, V., Pereira, A. P. D. S., Soares, F., & Esteves, J. S. (2017). Using a humanoid robot as the promoter of interaction with children in the context of educational games. *International Journal of Mechatronics and Applied Mechanics*, 1. Retrieved from https://pdfs.semanticscholar.org/b6e0/a9e38296176090c4e15d715ab739ce950a02.pdf

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innoenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71(2)*, 149-164. doi:10.1177/001440290507100202

*Hedgecock, J., Standen, P. J., Beer, C., Brown, D., & S. Stewart, D. (2014). Evaluating the role of a humanoid robot to support learning in children with profound and multiple disabilities. *Journal of Assistive Technologies*, *8*(3), 111-123. doi:10.1108/JAT-02-2014-0006

*Howard, A. M., & Park, H. W. (2014). Using tablet devices to engage children with disabilities in robotic educational activities. *Journal on Technology and Persons with Disabilities, 2,* 96-107. Retrieved from http://scholarworks.csun.edu/bitstream/handle/10211.3/133378/JTPD201412-p96-107.pdf?sequence=1

*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. doi:10.1109/TLT.2011.28

*Huijnen, C. A., Lexis, M. A., Jansens, R., & de Witte, L. P. (2016). Mapping robots to therapy and educational objectives for children with autism spectrum disorder. *Journal of Autism Development Disorders, 46,* 2100-2114. doi:10.1007/s10803-016-2740-6

*Jordan, K., King, M., Hellersteth, S., Wiren, A., & Mulligan, H. (2013). Feasibility of using a humanoid robot for enhancing attention and social skills in adolescents with autism spectrum disorder. *International Journal of Rehabilitation Research, 36*(3), 221–227. doi:10.1097/MRR.0b013e32835d0b43

*Jung, S. E., Lee, K., Cherniak, S., & Cho, E. (2019). Non-sequential learning in a robotics class: Insights from the engagement of a child with autism spectrum disorder. *Technology, Knowledge and Learning*, *25*(1), 63-81. doi:10.1007/s10758-018-9394-8

Jung, S. E., & Won, E. S. (2018). Systematic review of research trends in robotics education for young children. *Sustainability, 10*(4), 905. doi:10.3390/su10040905

*Karna-Lin, E., Pihlainen-Bednarik, K., Sutinen, E., & Virnes, M. (2006). Can robots teach? Preliminary results on educational robotics in special education. In *Proceedings of the Sixth International Conference on Advanced Learning Technologies* (pp. 319-321). Washington, DC: IEEE Computer Society.

Kim, T. Y. (2010). Sociocultural dynamics of ESL learning (de)motivation: An Activity Theory analysis of two adult Korean immigrants. *Canadian Modern Language Review, 67*(1), 91-122. doi:10.3138/cmlr.67.1.091

Kitchenham, B. (2004). *Procedures for performing systematic reviews*. Retrieved from http://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf

*Lee, H. W., & Hyun, E. J. (2015). The Intelligent robot contents for children with speech-language disorder. *Educational Technology & Society, 18*(3), 100–113.

*Lindsay, S., & Hounsell, K. G. (2016). Adapting a robotics program to enhance participation and interest in STEM among children with disabilities: A Pilot study. *Disability and Rehabilitation: Assistive Technology*, *12*(7), 694-704. doi:10.1080/17483107.2016.1229047

*Ludi, S., & Reichlmayr, T. (2011). The Use of robotics to promote computing to pre-college students with visual impairments. *ACM Transactions on Computing Education (TOCE)*, *11*(3), 1-20. doi:10.1145/2037276.2037284

Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. (n.d.). *Special need education*. Retrieved from http://www.mext.go.jp/en/policy/education/elsec/title02/detail02/1373858.htm

Moeller, J. D., Dattilo, J., & Rusch, F. (2015). Applying quality indicators to single-case research designs used in special education: A Systematic review. *Psychology in the Schools, 52*(2), 139-153. doi:10.1002/pits.21801

*Mohammad, S., Khaksar, S., Slade, B, Wallace, J., & Gurinder, K. (2019). Critical success factors for application of social robots in special developmental schools: Development, adoption, and implementation. *International Journal of Educational Management, 34*(4), 677-696. doi:10.1108/IJEM-08-2019-0304

*Park, E., & Kwon, S. J. (2016). I can teach them: The Ability of robot instructors to cognitive disabled children. *Journal of Psychological & Educational Research*, *24*(1), 101-114.

Pearson, S. (2009). Using activity theory to understand prospective teachers' attitudes to and construction of special educational needs and/or disabilities. *Teaching and teacher education*, *25*(4), 559-568.

*Pennington, R., Saadatzi, M. N., Welch, K. C., & Scott, R. (2014). Using robot-assisted instruction to teach students with intellectual disabilities to use personal narrative in text messages. *Journal of Special Education Technology, 29*(4), 49-58. doi:10.1177/016264341402900404

Perenboom, R. J. M., & Chorus, A. M. J. (2003). Measuring participation according to the International Classification of Functioning, Disability, and Health (ICF). *Disability and Rehabilitation, 25*(11-12), 577-587. doi:10.1080/0963828031000137081

Pearson, S., & Ralph, S. (2007). The Identity of SENCos: Insights through images. *Journal of Research in Special Educational Needs, 7*(1), 36–45.

*Pihlainen, K., Montero, C. S., & Kärnä, E. (2017). Fostering parental co-development of technology for children with special needs informal learning activities. *International Journal of Child-Computer Interaction, 11*, 19-27. doi:10.1016/j.ijcci.2016.10.010

*Saadatzi, M. N., Pennington, R. C., Welch, K. C., & Graham, J. H. (2018). Effects of a robot peer on the acquisition and observational learning of sight words in young adults with autism spectrum disorder. *Journal of Special Education Technology, 33*(4), 284-296. doi:10.1177/0162643418778506

*Van den Heuvel, R. J. F., Lexis, M. A. S., & de Witte, L. P. (2017a). Can the IROMEC robot support play in children with severe physical disabilities? A Pilot study. *International Journal of Rehabilitation Research, 40*(1), 53-59. doi:10.1097/MRR.0000000000000200

*Van den Heuvel, R. J. F., Lexis, M. A. S., & de Witte, L. P. (2017b). Robot ZORA in rehabilitation and special education for children with severe physical disabilities: A Pilot study. *International Journal of Rehabilitation Research, 40*(4), 353-359. doi:10.1097/MRR.0000000000000248

*Van den Heuvel, R. J. F., Lexis, M. A. S., & de Witte, L. P. (2020). ZORA robot based interventions to achieve therapeutic and educational goals in children with severe physical disabilities. *International Journal of Social Robotics*, *12*(2), 493-504. doi:10.1007/s12369-019-00578-z

*Wright, J. C., Knight, V. F., Barton, E. E., & Edward-Bowyer, M. (2019). Video prompting to teach robotics and coding to middle school students with Autism Spectrum Disorder. *Journal of Special Education Technology*. Advance Online Publication. doi:10.1177/0162643419890249

*Yuen, T., Mason, L. L., & Gomez, A. (2014). Collaborative robotics projects for adolescents with autism spectrum disorders. *Journal of Special Education Technology, 29*(1), 51-62. doi:10.1177/016264341402900104

Supplementary materials

Tables S1, S2, and S3 can be found on
https://drive.google.com/file/d/1jmjSQH4qodvD_czpuP18ehds_7him0zB/view?usp=sharing